Thinking Graphically: Extracting Local and Global Information

Raj M. Ratwani
(rratwani@gmu.edu)
George Mason University
Fairfax, VA 22030

J. Gregory Trafton
(trafton@itd.nrl.navy.mil)
NRL Code 5513
Washington DC, 20375

Deborah A. Boehm-Davis
(dbdavis@gmu.edu)
George Mason University
Fairfax, VA 22030

Abstract

This study investigates how information is extracted from a graph when different types of questions are asked. Although the process for extracting local information from simple graphs is understood quite well, the processes used to extract global information from more complex graphs are not as clear. In a series of two studies using verbal protocols and eye tracking, we compared responses to local and global questions. We replicated previous research on local questions, and show that people extract global information using a different set of cognitive processes.

Introduction

Several frameworks (Bertin, 1983; Kosslyn, 1989; Lohse, 1993; Pinker, 1990) and modifications to those frameworks (Peebles & Cheng, 2002; Carpenter & Shah, 1998; Trafton & Trickett, 2001) have been proposed to explain how information is extracted from graphs. These frameworks provide a broad set of cognitive operations that can be applied in different situations. Pinker (1990), for example, suggests a general task analysis that allows a "conceptual question" to be posed, a set of cognitive operations to be applied (e.g., relating information to long term memory via "graph schemas"), and a "conceptual message" to be extracted from the graph.

The focus of most of these models, theories, and experiments has been on the extraction of "local" information from simple and moderately complex graphs. For example, Pinker (1990) asked people to determine the answer to questions such as "What is the price of graphium in 1983?" Consequently, these kinds of local information extractions (also called read-offs) are quite well understood. In fact, local extractions can be described by a reasonably consistent set of steps that occur in a reasonably consistent order.

First, participants read a question to determine what information they are being asked to extract from the graph (e.g., What is the price of tin in 2001?). Parts of the question may be read multiple times (Peebles & Cheng, 2002). Next, the participant searches for the specific information on the graph, shifting from the axes to the main part of the graph and back again (Lohse, 1993; Kosslyn, 1989; Pinker, 1990; Carpenter & Shah, 1998). Once the information is found, multiple saccades occur between the main part of the graph and the legend in order to keep the information in memory (Carpenter & Shah, 1998; Trafton, Marshall, Mintz, & Trickett, 2002). Finally, the question itself is answered.

Less is known about what happens when people are asked to extract global or trend information from a graph. Empirical data suggest that global questions take longer and are more difficult to answer than local questions (Guthrie, Weber, & Kimmerly, 1993; Lohse, 1993). Specifically, Lohse (1993) found that the more difficult the question was, the longer it took to answer the question. Further, Guthrie et al. (1993) found that local questions elicited more explicit category-related extractions (i.e., reading off the axes on a bar-graph) and explicit read off information than did global questions while global questions elicited more general (global) abstractions.

The processes that might underlie these differences between local and global extractions have not been elaborated. Although Lohse (1993) showed that the most important determinant to reaction time was the number of cognitive operations needed to answer a specific question, he did not define global questions as needing a different set of cognitive operations. Other current models (Kosslyn, 1989; Pinker, 1990) do not differentiate between different types of questions. Thus, although they may be able to account for the results, they offer no predictions about the processes that people use to extract global information from graphs.

Our research goal was to show that there are qualitative differences between the way people answer global and local questions and that different questions activate different cognitive operations.

Experiment 1

The first experiment was designed to examine whether the type of question asked influences the cognitive processes used by individuals to answer those questions in the context of complex graphs. Thus, we chose to use graphs that were more complex than the graphs that have been used to date in this type of research. Our work uses choropleth graphs, which use different colors, shades of gray, or patterns to represent different quantities.

Method

Participants

The participants were ten George Mason University undergraduate psychology students who received course credit for their participation.

Materials

Four sets of choropleth graphs were created. Each set consisted of three to ten conceptually related graphs. For example, one set of three graphs showed the population for the years 1990, 1995, and 2000. Two sets of graphs were complex, containing 53 counties (see Figure 1 for an example).

The remaining two sets of graphs were less complex; each graph in those sets showed nine counties.

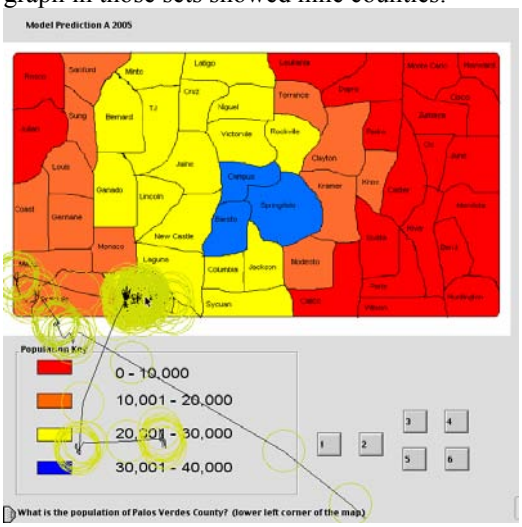

Figure 1. Graphs used in Experiments 1 and 2.

Four types of questions were generated for each set of maps: **describe** questions (which asked for a general description of what the graph represented); **global** questions (which required general trends or descriptions to be identified); **local** questions (which required straightforward single extractions from the graph); and **multi-search** questions (which asked for information that could only be obtained by searching the graph in multiple locations and had features of both local and global questions). Examples of each type of question are illustrated in Table 1.

Code	Example
Describe Question	Describe what is going on in this graph.
Global Question	What is the general trend of population growth in this graph?
Local Question	What is the population of Victorville county?
Multiple-Search Question	Which counties have the greatest populations?

Table 1. Examples of each question type.

Design

Participants were randomly assigned either to the local or global condition. Participants in both the global and local conditions were asked both global and local questions. Participants in the global condition answered global questions first while the local condition answered local questions first.

Procedure

In this experiment, all participants first saw a graph and then received an orientation question to allow them to become familiar with it. This orientation question asked the participant to "describe what was going on in the graph." Participants in the local condition then received a local question for that specific graph, and participants in the global condition then received either a global question or a multi-search question for that specific graph. This process continued for each of the graphs in the set. At the end of the set, all participants were asked to answer five to ten questions about those graphs. These questions were inference questions and were the same for both conditions.

Each graph was presented on a single sheet of paper. Participants were instructed to answer every question at their own pace. Participants were permitted to look back at any of the graphs needed. Each participant provided a talk-aloud protocol (Ericsson & Simon, 1994) as they examined the graphs and answered the questions. The participants' verbal protocols and the graphs they were examining were videotaped.

Coding Scheme

Transcriptions of the verbal protocols were coded prior to data analysis. The first step was to segment the protocols into individual utterances. Utterances were defined as a single thought and utterances that were not germane to the task at hand were coded as "off task" and eliminated from further analysis. Each remaining utterance was then coded as being an **aggregate read-off** (extracting conceptual information from the graph), a **specific read-off** (extracting individual information from the graph), explicit **search** (looking for a specific object or county), or **reasoning** (constructing a "story" of what is happening in the graph or making inferences about the data). A second independent coder coded 25% of the protocol data. Inter-rater reliability was calculated using Cohen's Kappa, kappa = .923, ($p<.001$), with inter-rater agreement at 94.2%. Table 2 shows examples of each type of utterance.

Code	Example
Aggregate Read-off	There is more blue on the graph, and less orange.
Specific Read-off	The population of Victorville county is 20,451 to 35,622
Search	Victorville, Victorville, Victorville, I don't see Victorville.
Reasoning	Since the outside seems to be the country area the center will grow.

Table 2. Examples of each utterance type.

Results and Discussion

Our two manipulations in Experiment 1 were question type (describe, local, global, multi-search) and condition (global, local). There were no significant differences between the global and local conditions in how they answered different types of questions (all $p > .05$); thus, we collapsed across condition for all analyses. Table 3 shows the percentage of each type of utterance for the entire experiment.

Utterance Type	Overall Frequency	Percentage
Aggregate Read Off	899	47.39
Specific Read Off	637	33.58
Search	145	7.644
Reasoning	216	11.39

Table 3: Frequency and Percentage of Different Utterances.

Our main goal was to show differences between types of questions and how participants answered these types of questions. Because search and reasoning accounted for a relatively small proportion of utterances, we focused on the number and type of extractions (specific read-off and aggregate read-off). To analyze these data, we normalized the raw frequencies by dividing the number of each extraction type by the number of questions that were asked.

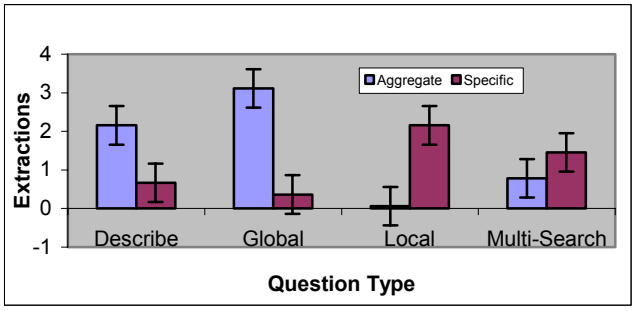

Figure 2. Average extractions per question

As Figure 2 suggests, participants extracted more aggregate information than specific information when answering describe questions, $F(1,9)=27.1$, MSE=.408, $p < .001$ and when answering global questions, $F(1,9) = 61.3$, MSE = .617, $p < .001$. Thus, when participants were describing the general trends of the graph or answering a global question, they extracted much more aggregate information from the graph than specific information.

In contrast, participants answering local questions extracted far more specific information from the graphs than aggregate information, $F(1,9) = 133.7$, MSE=.164, $p < .001$. Participants answered multi-search questions using roughly equivalent amounts of specific and aggregate information ($F(1,4)=5.2$, MSE=.216, $p = .08$), although there was a trend suggesting that they extracted a bit more specific information than aggregate information. As Figure 1 suggests, the interaction between question type and extraction type is highly significant, $F(1,9) = 153.8$, MSE=.382, $p < .001$.

Clearly, participants extracted differential types of data for different question types. Not surprisingly, for describe and global questions, participants extracted primarily aggregate information (i.e., the biggest counties are right next to each other), while for local and multi-search questions they extracted primarily local information.

Did they extract this information in different orders? To examine this issue, we calculated transition probabilities and created transition diagrams for each question type. To do this, we looked at the sequence of utterances in the verbal protocols. We then coded each pair of utterances (1-2; 2-3) by the type of utterance each pair represented (e.g., search followed by search, S-S, or search followed by specific read-off, S-SR). The total proportion of each type of transition was then calculated and diagrams constructed to illustrate those transition probabilities. The diagrams themselves include only those links that occurred more than 3% of the time for that question type.

We found that there was an overall difference in the pattern of transition probabilities as a function of question type, $\chi^2(15) = 217.3$, $p < .001$. Pairwise analyses with a Bonferroni adjustment showed that the describe questions are not significantly different from global questions, and local questions are not significantly different from multi-search questions, but all other comparisons were significant.

Thus, describe and global questions were answered in much the same way, and local and multi-search questions were answered in a similar way, but the manner in which global/describe questions and local/multi-search were answered was very different. Because the global and describe questions are so similar, we will illustrate the process differences using only the global questions. Similarly, because local and multi-search questions were so similar, we will only discuss the process used to answer local questions.

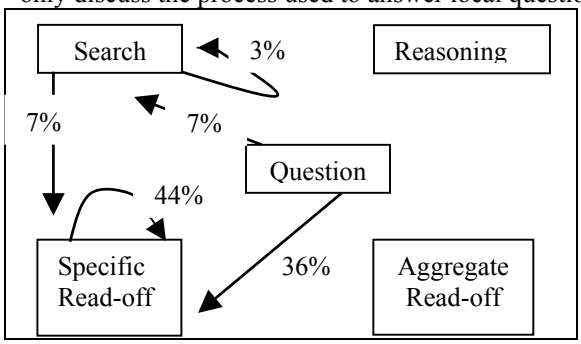

Figure 3. Local question transition diagram.

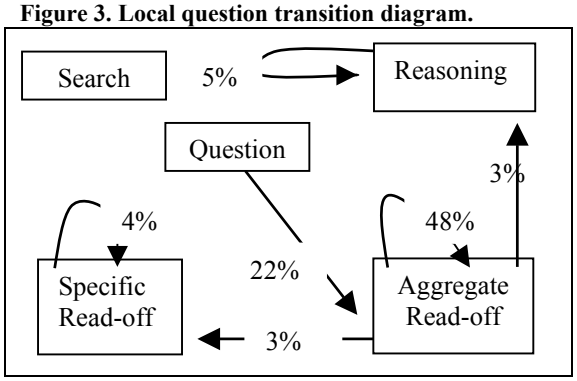

Figure 4. Global question transition diagram.

As Figures 3 and 4 suggest, participants answering local questions spent much of their time making specific read-offs and searching. In contrast, when participants answered global questions, they spent most of their time making aggregate read-offs and reasoning.

These data clearly show that there are large differences in the type and order of cognitive operations used: local questions elicit primarily search and specific read-offs while global questions elicit primarily reasoning and aggregate read-offs. At one level this is not a surprising result: global questions elicit global extractions and local questions elicit local extractions. However, it does show that current graph comprehension models and theories need to be updated to show sensitivity to the type of question asked as they do not predict these sorts of differences.

Experiment 2

In the first experiment, we clearly showed that there were high-level differences in the utterances that participants gave as they answered different types of questions. We also showed that the patterns of use of these utterance types differed as they answered different types of questions. This raises the question of how people might be visually examining the maps. These differences should surely translate into different ways of visually examining the maps.

The protocols from experiment one, theories of visual search, and previous work on graph comprehension suggest that when answering local questions, participants should visually search for the target, probably by systematically examining areas that catch the attention, then continuing on to other areas (McCarley, Wang, Kramer, Irwin, & Peterson, in press; Wolff, 1996). After finding the target, participants will probably saccade back and forth to the legend (Carpenter & Shah, 1998; Trafton et al., 2002) to read off the information, then answer the question.

Global questions will presumably show a different pattern of eye movements, but how those differences will be shown is not clear. Participants could systematically search county by county to understand differences at that level. Alternatively, participants could focus more on larger scale areas across the map. In this case, participants would spend far more effort on counties that are next to different colors ("edge" counties) to understand the size and shape of the different centroid areas (Lewandowsky, Herrmann, Behrens, Li, Pickle, & Jobe, 1993). Given the large qualitative and quantitative differences we found between global and local questions in Experiment 1, we believe that participants will visually examine maps by focusing on more counties that border another color and spend proportionally less effort focusing on the legend. This experiment was designed to use eye movement data to test these hypotheses.

Method

Participants

Twenty-one George Mason University undergraduate psychology students served as participants for course credit. One participant could not be calibrated on the eye tracker; his data were removed from all analyses.

Materials

The same sets of graphs used in the first experiment were used in the second experiment. For the second experiment, the materials (graphs and questions) were displayed on a computer screen. Graphs were shown in the center of the screen; questions (global, local or multiple search) were displayed at the bottom of the screen. To reduce the amount of time students took answering the questions, describe questions were eliminated from this study. Eye track data were collected using an LC Technologies Eyegaze System eye tracker operating at 60Hz (16.7 samples/second).

Design

The design was the same as Experiment 1, with 10 participants in the global condition and 10 participants in the local condition.

Procedure

The procedure was very similar to that used in Experiment 1; however, the use of the computer and eye tracker did necessitate some changes. The participants were seated at a comfortable distance from the monitor and used a chin rest. Participants first were calibrated on the eye tracker. Participants were then shown each map and the question(s) relevant for that map. The interface allowed participants to progress from map to map with a button-click and to look at maps they had previously viewed.

Coding Scheme

In these analyses, we examined a representative subset of the questions (two global questions and three local questions). Frequencies and transition diagrams were created by counting the number of gazes (via saccades) to different areas of the graph. The areas of the graph that were coded were: the legend, the title of the graph, and the main part of the graph itself.

If a participant gazed at the main part of the graph, it was coded in two additional ways: location of county and whether or not they read. For the location coding, if other counties of the same color surrounded the gazed-at county, it was coded as an "inner" county. If the county was on an edge between one or more different colored counties, it was coded as an "edge." If the county was on the outside border of the map, it was coded as a "border." For the read coding, if the participant read the name of the county, it was coded as "read." If the participant looked at a county but did not read the county's name, it was coded as "not read." Figure 5 shows an annotated example of each different coding type.

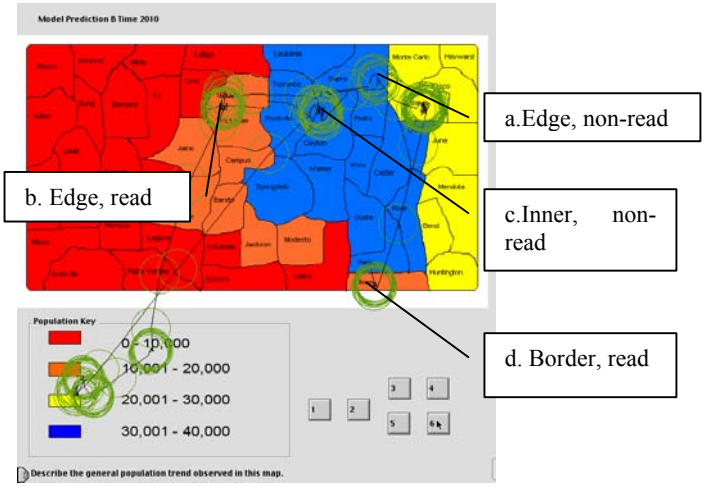

Figure 5. Sample eye-movements with examples of each gaze type.

Results and Discussion

Experiment 2 was designed to expand on the process differences found in Experiment 1. Specifically, we wanted to examine if participants' eye movements differed across global and local questions, and, if so, how they differed. In general, because global questions are more complex (Guthrie et al., 1993; Lohse, 1993) and elicited many more gazes and utterances, we perform our statistics on percentages, though we also present raw frequencies.

For local questions, we expected to find an initial search for the county names with more reading than non-reading, finding of the target, and several back and forth saccades between the legend and the target. For global questions, we expected to find participants gazing more frequently at edge counties, and cycling back and forth between edges. We also expected to find proportionally fewer legend gazes overall when answering a global question than when answering a local question.

	Local	Global
Edge	3.8 (39%)	17.4 (56%)
Inner	1.3 (13%)	9.8 (26%)
Border	1.7 (17%)	1.0 (3%)
Legend	3.1 (31%)	4.1 (15%)
Read	5.8 (85%)	12.2 (46%)
Non-Read	1.0 (15%)	14.1 (54%)

Table 4. Kinds of counties examined, whether or not they read the county names, and number of times the legend was examined (averaged by question).

In general, this is exactly what we found (see Table 4). Participants read county names more often than not in the local condition, but reading behavior did not differ in the global condition, $\chi^2(1) = 32.0$, $p < .0001$, (Bonferonni adjusted χ^2 significant at $p < .001$ for local questions, $p > .10$ for global questions).

We also found that when participants answered local and global questions, they differed in the number and type of counties they examined, $\chi^2(3) = 22.7$, $p < .0001$. The main source of this difference between question types seems to be that when participants answered local questions they focused more on the legend and less on the edge counties, while global questions elicited more edge gazes and fewer legend gazes, $\chi^2(1) = 7.8$, $p < .01$.

Figures 1 and 6 show examples of participants answering a local question and a global question, respectively. Notice that Figure 6 shows a participant focusing primarily on edge counties, tracing the boundaries of the color divider. In contrast, Figure 1 shows the participant searching for the target, finding it (obscured in the figure by the eye-track), and then sacadding down to the legend to read off the value.

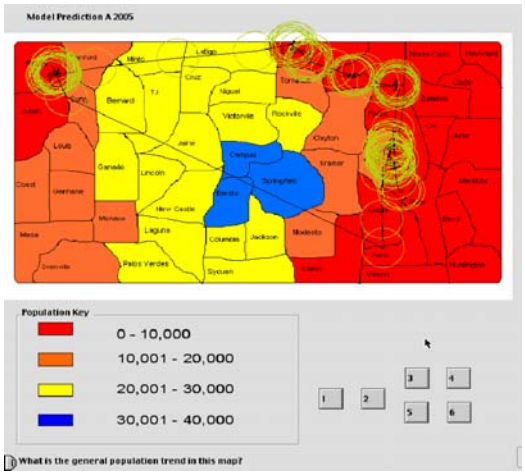

Figure 6. Sample eye-movements from a participant answering a global question.

Comparing the number of edge gazes to the other gaze types also shows an interesting pattern: people in both conditions examined edge counties more often than other county types (Bonferonni adjusted χ^2 significant, $p < .008$). When answering a global question, this makes sense. However, when answering local questions, participants showed the same pattern. This could be due to participants having their eyes drawn to differential colors. It also could be that participants are searching for county names from edge to edge, using the same processes used when answering global questions. To investigate these possibilities, we created transition graphs (see Figures 7 and 8).

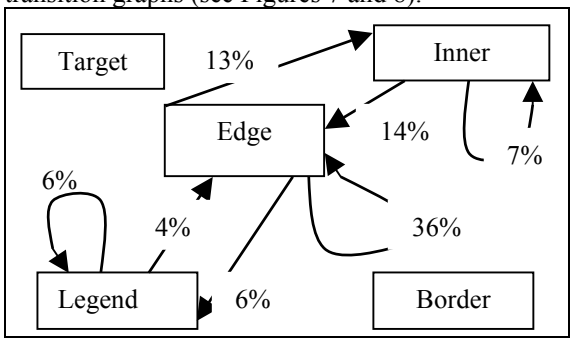

Figure 7. Global transition graph from Experiment 2.

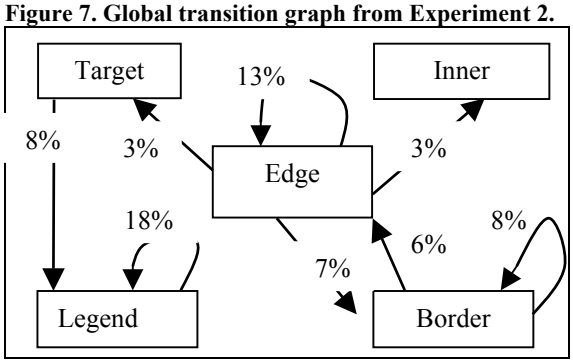

Figure 8. Local transition graph from Experiment 2.

As Figures 7 and 8 suggest, answering global and local questions engage different types and orders of eye movements. Importantly, when answering local questions, looking from edge to edge occurs less frequently than when

answering a global question, $\chi^2(1) = 10.8$, $p < .05$. Thus, consistent with our earlier analysis, when participants answer a global question, they look from edge to edge to understand the size and shape of the centroid. When answering a local question, participants search, find the target, and then examine the legend.

One part of our earlier analysis did not, however, hold up. Previous research has shown that when participants need to extract information from a legend, they saccade back and forth between the main graph and the legend (Carpenter & Shah, 1998; Trafton et al., 2002). In this study, we found that when participants answered local questions, they found the target and immediately went to the legend (see Figure 1). However, after a single examination of the legend, they answered the question. This result is somewhat odd because it seems to contradict a robust, replicable effect. We believe that there are two main possibilities for this finding. First, it could be that because participants saw these graphs over and over again, they became more familiar with the legend and essentially memorized them, needing only a reminder gaze at the legend. Alternatively, our legend was rather bigger than that used in previous studies, and that may have affected the overall gaze performance in some way.

General Discussion

Choropleth graphs were used in these experiments because of their complexity. This complex graph type allows us to generalize to complex representations such as meteorological graphs and scientific visualizations. Experiment 1 showed that there were major process differences in how people answer local and global questions. First, local questions elicited the standard search → find → answer behavior that has been found in previous studies. However, contrary to other graph comprehension theories, we showed that the cognitive steps that are followed to answer a global question are quite different. In general, global questions were answered by a series of aggregate read-offs.

Experiment 2 expanded on this finding by showing big differences in how people visually inspected graphs when asked local and global questions. Local questions showed a search (read) → find → legend → answer behavior, while global questions showed a trace-edges (don't read) → answer behavior. We believe that this edge-tracing behavior allows participants to understand the general shape of large map features, which in turn allows them to describe what is occurring in the graph at a high level without becoming overly concerned with individual data elements.

What are the implications for current theories of graph comprehension? One obvious implication is that they should not assume that different questions use the same mental operations. Using different operations to account for different question types should allow better theories and models to be built. Second, even though several current models suggest sacadding back and forth between the main graph and the legend, this does not seem to be true in all cases.

Finally, how people go about visually inspecting a graph is much more complex than what we have described here. It seems that people use several heuristics to search and a completely different set of heuristics to explore. This search/explore methodology is not accounted for in any theories of graph comprehension, and it is not immediately obvious how to easily integrate this information into such theories.

Acknowledgements

This research was supported in part by grant 55-7850-00 to the second author from ONR and by George Mason University. We thank Mike Schoelles for designing the interface of the eye tracker and Brandon Beltz for IRR coding.

References

Bertin, J. (1983). Semiology of graphs. Madison, WI: University of Wisconsin Press.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. Journal of Experimental Psychology: Applied, 4 (2), 75{100.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data (Revised edition). Cambridge, MA: MIT Press.

Guthrie, J., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. Contemporary Educational Psychology, 18 186-221.

Kosslyn, S. M. (1989). Understanding charts and graphs. Applied Cognitive Psychology, 3, 185-226.

Lewandowsly, S., Herrmann, D.J., Behrens, J.T., Li, S-C, Pickle, L., Jobe, J.B., (1993). Perception of clusters in statistical maps. *Applied Cognitive Psychology,*7, 533-551.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. Human Computer Interaction, 8, 353-388.

McCarley, J. S. , Wang, R. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (in press) How Much Memory Does Oculomotor Search Have? Psychological Science.

Peebles, D., & Cheng, P. C.-H. (2002). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research,* 3, 77-86.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future of testing, (pp. 73-126). Hillsdale, NJ:

Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). Extracting explicit and implict information from complex visualizations. In H. Narayanan (Ed.), Diagramatic Representation and Inference (pp. 206-220). Berlin: Springer-Verlag.

Trafton, J. G., & Trickett, S. B. (2001). A new model of graph and visualization usage, *The proceedings of the twenty third annual conference of the cognitive science society* . Mahwah, NJ: Erlbaum.

Wolfe J.M. (1996). Extending guided search: Why guided search needs a preattentive "item map." In: A.F. Kramer, M.G.H. Coles, & G.D. Logan (Eds), Converging operations in the study of visual attention. (pp. 247-270) Washington, DC: American Psychological Association.