Journal of Experimental Psychology: Applied
2008, Vol. 14, No. 1, 36–49

1076-898X/08/$12.00 DOI: 10.1037/1076-898X.14.1.36

Thinking Graphically: Connecting Vision and Cognition During Graph Comprehension

Raj M. Ratwani
George Mason University

J. Gregory Trafton
Naval Research Laboratory

Deborah A. Boehm-Davis
George Mason University

Task analytic theories of graph comprehension account for the perceptual and conceptual processes required to extract specific information from graphs. Comparatively, the processes underlying information integration have received less attention. We propose a new framework for information integration that highlights visual integration and cognitive integration. During visual integration, pattern recognition processes are used to form visual clusters of information; these visual clusters are then used to reason about the graph during cognitive integration. In 3 experiments, the processes required to extract specific information and to integrate information were examined by collecting verbal protocol and eye movement data. Results supported the task analytic theories for specific information extraction and the processes of visual and cognitive integration for integrative questions. Further, the integrative processes scaled up as graph complexity increased, highlighting the importance of these processes for integration in more complex graphs. Finally, based on this framework, design principles to improve both visual and cognitive integration are described.

Keywords: graph comprehension, diagrammatic reasoning, eye tracking

Supplemental materials: http://dx.doi.org/10.1037/1076-898X.14.1.36.supp

Imagine a health scientist attempting to contain the flu epidemic. The scientist may study a population density graph (similar to Figure 1) depicting the number of flu cases by county. To allocate the limited flu vaccine, the scientist may want to know the specific number of cases in a particular county or the general trend of infected cases. What cognitive and perceptual processes are used to extract these different types of information from a graph?

Specific information extraction involves searching for a specific data point and determining the quantity represented by the data point. Integrating information, which is required to determine the trend in a graph, demands the extraction of multiple data points and interpretive processes to combine this information into a coherent representation. The goal of this article is to examine the cognitive and perceptual processes used to extract specific information and to integrate information, with an emphasis on the integration process. We elaborate on existing frameworks of integration and provide guidelines to facilitate integration in graphs.

Specific Information Extraction

Most theories of graph comprehension have focused on the cognitive and perceptual processes underlying specific information extraction from graphs. Specifically, task analytic theories (Kosslyn, 1989; Lohse, 1993; Pinker, 1990) provide process description of specific information extraction with a focus on relatively simple graph types (i.e., bar and line graphs). These theories are remarkably consistent, suggesting the following stages of processing: (a) pattern recognition occurs, whereby graph readers encode the visual array and identify visual features of the graphical pattern; (b) conceptual relations are determined, giving rise to the quantitative meaning of the visual features; and (c) referents of the graph are related to the encoded visual features.

Several empirical studies have provided support for the task analytic theories' account of specific information extraction by examining the specific processes that comprise the general stages outlined previously (Carpenter & Shah, 1998; Lohse, 1993; Peebles & Cheng, 2003). First, parts of the question may be read multiple times (Peebles & Cheng, 2003). Next, the participant searches for the specific information on the graph (Stage 1), shifting from the axes to the main part of the graph and back again (Stage 2; Carpenter & Shah, 1998; Lohse, 1993; Pinker, 1990).

Raj M. Ratwani and Deborah A. Boehm-Davis, Department of Psychology, George Mason University; J. Gregory Trafton, Naval Research Laboratory.

This research was supported in part by Grant 55-7850-00 from the Office of Naval Research to J. Gregory Trafton. We thank Mike Schoelles for his help in programming the experiments. We thank Brandon Beltz, Cara Stitzlein, and David Cades for their help with coding. Parts of Experiment 1 of this article were previously published in Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2003). Thinking graphically: Extracting local and global information. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Boston, MA: Erlbaum.

Correspondence concerning this article should be addressed to Raj M. Ratwani, Department of Psychology, George Mason University, MSN 3F5, Fairfax, VA 22030-4444. E-mail: rratwani@gmu.edu

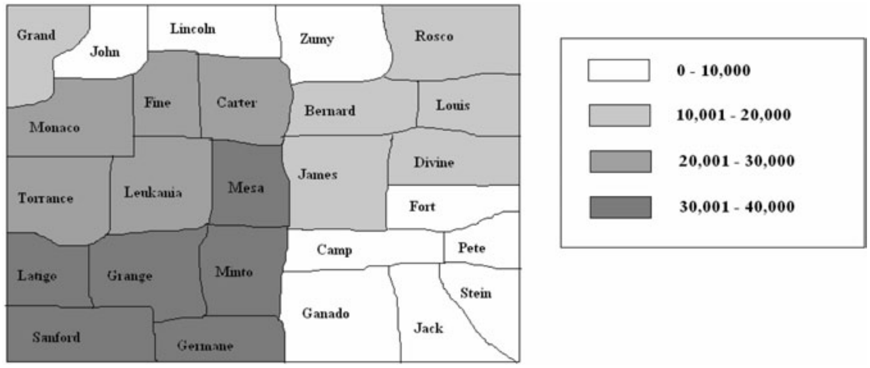

Figure 1. General example of the type of graph used in Experiments 1–3; the different shades of gray were actually different colors. The actual figure can be viewed at http://dx.doi.org/10.1037/1076-898X.14.1.36.supp.

Once the information is found, multiple saccades occur between the main part of the graph and the legend to keep the information in memory (Stage 3; Carpenter & Shah, 1998; Trafton, Marshall, Mintz, & Trickett, 2002). Finally, the question itself is answered.

The perceptual processes underlying specific information extraction have been closely examined as well. Cleveland and McGill (1984) identified a set of elementary perceptual tasks that are carried out when specific information is extracted from graphs. These perceptual tasks have been empirically examined and have resulted in a hierarchical ordering of graph types based on accuracy of information extraction. Following this work, Simkin and Hastie (1987) showed an interaction between the perceptual decoding of the graph and the judgment task the graph reader was attempting to accomplish. Simkin and Hastie suggested several elementary processes (e.g., projection and anchoring), which are used to make comparisons and proportion judgments. Thus, both the cognitive and perceptual processes underlying specific information extraction are accounted for by the task analytic theories and subsequent empirical studies.

Integration of Information

Historically, task analytic theories focused on the extraction of specific data points, thus these theories have not been applied with much success to graph integration. For example, Carswell (1992) evaluated the predictions of a basic task model of graphical perception, primarily based on Cleveland and McGill (1984), and found task models were more successful at accounting for specific extraction tasks than integration tasks. In the last 15 years, there have been three notable theories that go beyond pure information extraction to integration.

Carpenter and Shah (1998) extended the task analytic theories to account for integration by suggesting the same stages of processing used for specific information extraction occur for integration; however, multiple cycles of processing are required. This integrative framework included a pattern recognition stage and two interpretative stages. The pattern recognition stage leads to the encoding of a visual pattern by forming a visual chunk. The interpretive stages translate the pattern into its quantitative and qualitative interpretation and relate this information to the referents in the graph. These processes are repeated in a cyclical fashion for each visual chunk in the graph, with each cycle interpreting a single chunk.

To demonstrate this cyclical process, Carpenter and Shah (1998) examined graph readers' transitions between regions of the graph (e.g., the graphical pattern, axes, legend, title, etc.). Graph readers' fixations cycled between the different regions for each of the visual chunks represented in the graph, suggesting that graph readers cycled between different stages of processing. As graph complexity increased (i.e., the number of unique visual chunks), the number of transitions between regions of the graph increased, suggesting that a single processing cycle was required for each chunk in the graph. Using this high-level transition analysis, Carpenter and Shah extended the task analytic theories to account for some integration processes.

Gillian and Lewis (1994) proposed a model of graphical perception called mixed arithmetic-perceptual (MA-P). Gillian and Lewis suggested that the processes used to interpret a graph are task specific. MA-P has five stages of processing with several similarities to the task analytic theories; however, MA-P contains an explicit spatial component. For specific extraction tasks, the processes are similar to those of the task analytic theories; however, for integration the MA-P model relies on the identification of spatial relations (Gillian, 1995). This suggests that spatial processing is also required to account for information integration.

Wickens and Carswell (1995) have proposed the proximity compatibility principle, which suggests that information that needs to be integrated should be close in perceptual proximity. Perceptual proximity can take the form of being spatially proximate as well as being perceptually similar (e.g., sharing the same color coding). Thus, this principle also suggests that spatial processes may be required for integration. There are several manipulations to increase perceptual proximity, many of which follow from Gestalt laws of perceptual organization (for more on graphical integration, see Wickens and Hollands, 2000). An increase in perceptual proximity facilitates integration by reducing search cost and working memory load.

These different theories addressing integration illustrate several important aspects of the integration process. Carpenter and Shah (1998) showed that multiple cycles of forming a visual chunk and relating this information to the referents are needed. Gillian and Lewis (1994) and Wickens and Carswell (1995) showed the im-

portance of a spatial component. This article takes these findings as a starting point and proposes a novel process theory of integration.

We suggest that the integration process is more complex than any of these accounts alone. Integration during graph comprehension has at least two primary components: visual integration and cognitive integration. First, visual integration must occur between similar data points leading to visual clusters. Similarity will usually be based on perceptual features (e.g., color) but may also be based on semantic or knowledge-level features (e.g., areas that are semantically related), spatial features (e.g., proximity), or other salient features. These individual data points will be visually integrated to form higher order clusters of information. These visual clusters are similar to Liu and Wickens' (1992) concept of visual chunks of information. Second, once aggregate visual clusters have been formed through pattern recognition processes and related to the referents, integration must occur between visual clusters that are different in some way via a comparison–contrast mechanism. Thus, visual integration results in objectlike visual clusters that can be directly compared to form a coherent representation. This process of visual and cognitive integration is iterative and scales with the complexity of the graph: As graph complexity increases, more cycles of visual cluster formation and comparison are required.

This model of integration combines and further specifies the previous models substantially. The visual integration component, which focuses on explicit pattern recognition processes, provides detail at a lower level than Carpenter and Shah (1998). The cognitive integration component is a novel process of direct comparisons of visual clusters and differs from Carpenter and Shah's model. This model also describes the spatial component of integration more explicitly than Gillian and Lewis (1994) and Wickens and Carswell (1995) and further illustrates the importance of spatial processing, as other authors have claimed (Trafton et al., 2000; Trickett & Trafton, 2006).

Current Study

We examined graph readers' cognitive and perceptual processes as they answered specific information extraction and integration questions. We expect to find general support for the task analytic theories' account of how graph readers extract specific information. For integration questions, we expect to find general support for the multiple cycles of processing suggested by Carpenter and Shah (1998). In addition to this cyclical process, we expect to find evidence for visual cluster formation and comparison, and these processes should correspond to performance on the task.

Specifically, the visual integration process should manifest itself as explicit fixations defining distinct visual clusters of information. These visual clusters can then be explicitly compared to each other to develop a coherent representation of the graph. This cognitive integration can be shown in many ways; in this study we operationally define this process as comparisons (verbal or perceptual) between visual clusters. Visual and cognitive integration should scale up with complexity: More complex graphs (i.e., a greater number of distinct visual clusters) should elicit more explicit cluster formation and explicit comparisons of these clusters.

To explicitly examine the visual and cognitive integration processes, choropleth graphs were used (see Figure 1). These graphs use color or shading of regions to represent magnitude (Lewandowsky & Behrens, 1999) and are representative of a class of spatial color-coded graphs including meteorological and geological graphs as well as visualizations used in oceanography and several other scientific domains. Choropleth graphs were used because of the large graphical pattern with several distinct regions, which makes them conducive to examining the perceptual processes underlying visual and cognitive integration. Further, these graphs are more complex than traditional graphs used in graph comprehension studies, representing between 10 and 70 data points as compared to the 3 to 6 data points of most studies (Lohse, 1993; Pinker, 1990).

In Experiment 1, we examined graph readers' verbal protocols as they answered specific extraction and integration questions; the focus was on the pattern of processes at the verbal level.

In Experiment 2, the perceptual processes underlying specific information extraction and integration were examined. Experiment 3 focused solely on integration questions. Finally, in the General Discussion section, guidelines for designing graphs to facilitate integration are described.

Experiment 1

In Experiment 1, participants performed a verbal protocol as they were asked specific extraction and integration questions; the focus was on the pattern of processes as they answered these questions. We first sought to find evidence for the single cycle and multiple cycle processes for specific information extraction and integration, respectively. It is possible that the cyclical pattern of integration as suggested by Carpenter and Shah (1998) is limited to certain graph types; there have not been many empirical studies examining this multiple cycle process.

When extracting specific information, we expected to find the single cycle process of searching, extracting information, and finally answering the question. This should be apparent in the utterances from the verbal protocol data. Several empirical papers have shown support for these processes (Guthrie, Weber, & Kimmerly, 1993; Kosslyn, 1989; Lohse, 1993; Pinker, 1990; Shah & Hoeffner, 2002); we expect to find support as well.

In contrast, information integration (e.g., determining the trend) should elicit a different pattern of processes. The verbal protocol data should provide evidence for multiple cycles of processing as suggested by Carpenter and Shah (1998): Graph readers should form visual clusters (i.e., refer to different spatial regions of the graph), interpret these clusters in relation to the referents, and build a representation by cycling through these stages multiple times. Further, we sought evidence for a cognitive integration component; graph readers should explicitly compare different visual clusters (spatial regions) in the graph.

Method

Participants. Ten George Mason University undergraduate psychology students (six women and four men) participated for course credit.

Materials. Four sets of choropleth graphs were created, each contained 3 to 10 conceptually related graphs. For example, one set contained three graphs showing the population for the years 1990, 1995, and 2000. Two sets of graphs were complex, each

containing 53 counties (see Figure 1), and two sets of graphs were less complex containing nine counties.

Four types of questions were generated for each set of graphs: describe questions (asked for a general description of what the graph represented), integration questions (required general trends to be identified), specific extraction questions (required single extractions from the graph), and multiple search questions (multiple specific extraction questions). Because the focus of this study is on specific extraction and integration, we do not discuss the multiple search questions.

Design. A counterbalanced within-participants design was used. Participants first received the describe question to orient them with the graph (these data were not analyzed). Half of the participants then received the integration questions followed by specific extraction questions and the other half of the participants received the reverse order.

Procedure. Each graph was presented on a single sheet of paper with the questions written below the graph. Participants were instructed to answer every question at their own pace and were permitted to look back at any of the graphs as needed. Each participant provided a talk-aloud protocol (Ericsson & Simon, 1993) as they examined the graphs and answered the questions. The participants' verbal protocols and the graphs they examined were videotaped.

Coding scheme. Transcriptions of the verbal protocols were coded prior to data analysis. The protocols were segmented into individual utterances. Utterances were defined as a complete thought, and utterances that were not germane to the task were eliminated from further analysis (see Trickett and Trafton, 2007, for a full description of this process). Each remaining utterance was then coded in the following ways: *qualitative extraction* (extracting general conceptual information from the graph), *quantitative extraction* (extracting specific quantitative information from the graph), explicit *search* (looking for a specific object or county), *reasoning* (constructing a "story" of what was happening in the graph or making inferences that went beyond the data), or *cognitive integration* (making comparisons or forming relationships with the information extracted from the graph). A second independent coder coded 25% of the protocol data. Interrater reliability was calculated using Cohen's kappa, $\kappa = .91$, $p < .001$, with interrater agreement at 93.1%. See Appendix A for coding details.

Results and Discussion

Types of extractions. Raw frequencies were normalized by dividing the number of each extraction type by the number of questions that were asked (see Table 1). A repeated measures analysis of variance (ANOVA) was used to examine question type (specific extraction, integration) and extraction type (quantitative, qualitative). The main effect of question type was significant, $F(1, 9) = 8.7$, $MSE = .14$, $p < .05$, $\eta^2 = .5$, suggesting that participants extracted different types of information for specific questions as compared to integration questions. The main effect for type of extraction was marginal, $F(1, 9) = 3.9$, $MSE = .05$, $p = .08$, $\eta^2 = .3$. The interaction between question type and extraction type was significant, $F(1, 9) = 211.6$, $MSE = .05$, $p < .001$, $\eta^2 = .96$. To explore this interaction, we performed multiple Tukey's honestly significant difference (HSD) post hoc analyses. A greater number of quantitative extractions were made when specific extraction questions were answered ($M = 1.10$, $SD = 0.05$) as compared to integration questions ($M = 0.1$, $SD = 0.2$), $p < .01$. A greater number of qualitative extractions were made when integration questions were answered ($M = 1.5$, $SD = 0.5$) as compared to specific extraction questions ($M = 0.3$, $SD = 0.3$), $p < .01$. This clearly shows that graph readers extracted different types of information depending on the question type.

Transition diagrams. To examine the cycles of processes that occurred during specific information extraction and integration, we calculated transition probabilities and created one deep transition diagram for each question type. To do this, we looked at the sequence of utterances in the verbal protocols and coded each pair of utterances (1st utterance to 2nd utterance, 2nd utterance to 3rd utterance, and so on) by the type of utterance each pair represented (e.g., search followed by search, or search followed by quantitative extraction). A percent of each type of transition was then calculated by taking the proportion of each transition type relative to all transitions. Diagrams were constructed to illustrate these transition

Table 1
Frequency (Percentages in Parentheses) of Each Utterance Type by Question Type

Utterance category	Specific questions	Integration questions
Qualitative extraction	6 (1.8)	158 (47.3)
Quantitative extraction	283 (86.3)	34 (10.2)
Search	39 (11.9)	0
Reasoning	0	33 (9.9)
Cognitive integration	0	109 (32.6)

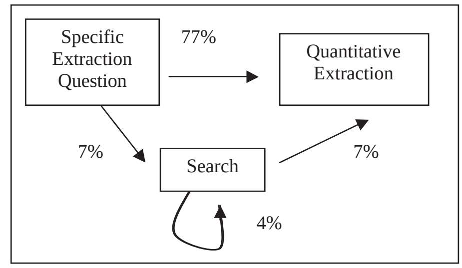

Figure 2. Transition diagrams for specific extraction questions.

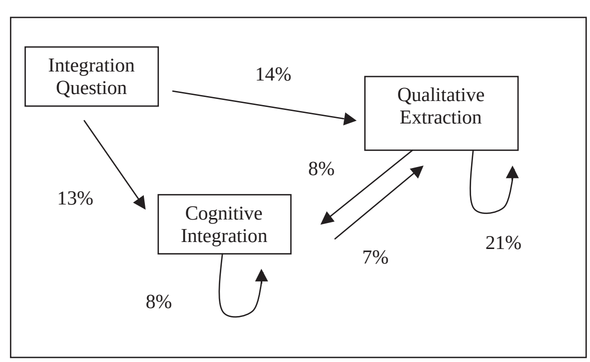

Figure 3. Transition diagram for integration questions.

probabilities; only those links that occurred 4% or more of the time are represented.

When graph readers extracted specific information, there was evidence of the single cycle process, as Figure 2 shows. The pattern of processes is in agreement with the task analytic theories; most of the time graph readers read the question and directly extracted quantitative information in one cycle. Occasionally they verbalized the search process and then extracted the information directly. Notice that search was the only repetitive process: Participants would occasionally make multiple utterances about the status of their search for the target.

In contrast, as shown in Figure 3, several cycles of processing were needed to answer the integration questions. This cyclical nature of processing was evident in several ways. First, cognitive integration utterances frequently followed other cognitive integration utterances. Second, cognitive integration and qualitative extraction cycled between each other. Finally, qualitative extractions frequently followed other qualitative extractions. When integrating information, the pattern of processes included cognitive integration: Graph readers made explicit comparisons between different areas of the graph as opposed to strictly making qualitative extractions. Statistical comparisons of the pattern of processes in Figures 2 and 3 were not performed, because the specific processes that occurred for each question type were entirely different from each other . Nonetheless, a visual comparison of Figures 2 and 3 clearly shows differences that will be explored in later analyses.

Although Experiment 2 illustrated multiple cycles of processing with cognitive integration for integration questions, the specific perceptual processes remain unclear. In particular, how were graph readers extracting qualitative information when answering integration questions and what perceptual processes underlie the process of cognitive integration? The overarching goal of Experiment 2 was to understand the perceptual processes that underlie the differences in specific information extraction and integration with a focus on the integration process.

Experiment 2

In Experiment 2, we collected eye movement and verbal protocol data as participants answered specific extraction and integration questions. First, we directly compared the perceptual processes during the pattern recognition stage as participants answered these questions. Second, we focused on the integration process and sought to tie the eye movement data to graph readers' verbal responses to understand how visual integration and cognitive integration give rise to a coherent representation of the graph.

Perceptual Processes During the Pattern Recognition Stage

To understand the perceptual processes underlying the differences in the type of information extracted by question type found in Experiment 1, we examined the location of graph readers' fixations on the graphical pattern. For specific information extraction, researchers have shown that graph readers examine specific locations to find the target and extract the value associated with that target point (Trafton et al., 2002). Thus, we expect graph readers' fixations to be concentrated on the inside of counties, reflecting a read and search process.

The perceptual processes underlying information integration are not as clear. Graph readers may focus on specific counties, just as with specific information extraction, and mentally combine this information in an aggregate manner resulting in qualitative responses. Alternatively, visual integration may be used during the pattern recognition stage. Visual integration involves the explicit formation of visual clusters of information. In the choropleth graphs used here, groups of same-colored counties may constitute a visual cluster (Brewer & Pickle, 2002; Herrmann & Pickle, 1996; Lewandowsky et al., 1993). Even though graph readers may form visual clusters in several ways, there is some evidence that fixating on distinguishable boundaries allows for segmentation into unitary objects (Bravo & Farid, 2002; Schyns & Oliva, 1994). Thus, we have used the explicit fixation to the boundaries of groups of same-colored counties as a measure of visual cluster formation.

These different hypotheses about integration lead to two diverging predictions in regard to the location of fixations. If graph readers are paying attention to individual counties, there should be no difference in fixation locations between specific extraction and integration questions. However, if visual integration is used, there should be a greater number of fixations to the boundaries of groups of same-colored counties; we term these fixations *cluster-boundary* fixations. Further, there should be a pattern to these cluster-boundary fixations such that they are in service of forming a quantifiable number of visual clusters.

Visual Clusters and Cognitive Integration

Next, we focused on tying the pattern recognition processes to cognitive integration. Our model of integration suggests two critical processes: (a) visual clusters are formed during the pattern recognition stage (visual integration); and (b) these visual clusters, in addition to being related to their referents, are directly compared to each other (cognitive integration). As graphs increase in complexity, the visual and cognitive integration processes should scale up as well. Although Experiment 1 provided some support for these processes, in Experiment 2 eye movement and verbal protocol data were collected to provide additional support for visual and cognitive integration by focusing on three critical points.

First, if the visual clusters formed during visual integration are used to reason about the graph during cognitive integration, the number of visual clusters found from eye tracking should relate to the number of verbal clusters in the verbal data. Second, after forming visual clusters, graph readers should relate these clusters not only to the referents of the graph but also to other visual clusters. Finally, if integration depends on comparing visual clusters to each other, then performance on the task, measured by quality of answer, should relate to these comparisons.

Method

Participants. Seventeen George Mason University undergraduate psychology students (10 women and 7 men) participated in Experiment 2 for course credit.

Materials. Choropleth graphs that displayed the populations of 32 counties were used in Experiment 2; each county was marked by a single unique letter or number positioned in the center of the county to allow for more accurate eye tracking. Each county differed in size; the smallest county subtended 2.4° of visual angle,

whereas the largest county subtended 3° of visual angle. A total of 14 different maps were generated; each map was 30 × 34 cm. Seven of the maps had three clusters, and seven of the maps had seven clusters. The coding scheme on the legend was in color and was the same for all of the graphs; this scheme did not follow any logical order. For each graph, participants were asked a specific extraction and integration question. The specific extraction question changed for every graph; the integration question remained the same and always asked the participant to determine the trend in the graph. The maps and questions were randomized for each participant. Eye movement data were collected using an LC Technologies Eyegaze Analysis System (LC Technologies, McLean, VA) heads-free eye tracker operating at 60 Hz (16.7 samples/s) with gaze position accuracy of less than 0.5° of error. The eye tracker used corneal reflection to record eye movements; a chin rest was used to reduce recalibration. The eye tracker was run from a single desktop personal computer running Windows 2000.

Design. The complexity of the graph (three cluster vs. seven cluster) and the type of question being asked (specific extraction vs. integration) were examined in Experiment 2 using a within-participants design. Each participant answered a specific extraction and integration question for each graph on independent trials. Thus, each participant viewed a total of 28 graphs and answered 14 questions of each type. The order of graph presentation was randomized.

Procedure. Participants were seated approximately 46 cm from the monitor and placed their chin on a chin rest for added head stability; the eye tracker was then calibrated. Each participant was instructed to read each question out loud and answer each question out loud as they examined each graph; the eye track data and the verbal responses occurred concurrently. Pilot testing showed that verbal responses did not disrupt the recording of eye movements; this was determined by having pilot participants look at predefined locations while talking and measuring their deviation from the predefined locations. Each question was displayed in the middle of a blank screen. The participant used the mouse to move on to the next screen, which displayed the graph and legend. The experiment was self-paced, and the participant could not return to view the screen that displayed the question for each particular graph; thus the participant had to remember the question for each graph (cf. Peebles & Cheng, 2003).

Coding scheme. A fixation was defined as a minimum of five eye samples (~100 ms) within 10 pixels (approx 2° of visual angle) of each other, calculated in Euclidian distance. The center of gravity of the fixation was based on taking the average of the Cartesian coordinates of all included eye samples. The location of each fixation was coded relative to the actual clusters of same-colored counties in the graph and was coded as either a cluster-boundary fixation or an inner fixation (see Figure 4). A cluster-boundary fixation was directed to the boundary between clusters of different colored counties. An inner fixation was completely in one color region, thus an inner fixation may have been to the junction of two same-colored counties or completely within a single county. Cluster-boundary and inner fixations were coded to test the hypothesis that visual clusters were formed by fixating on the boundaries of clusters as opposed to being formed by fixating on individual counties within the clusters. Each of the specific areas of the graph was defined as an area of interest to analyze the eye track data.

We also implemented a coding scheme to calculate the actual number of visual clusters formed in each graph. A sequence of fixations was defined as forming a visual cluster if it met one of two criteria. First, a consecutive sequence of at least two fixations to opposite edges of boundaries of a single cluster of same-colored

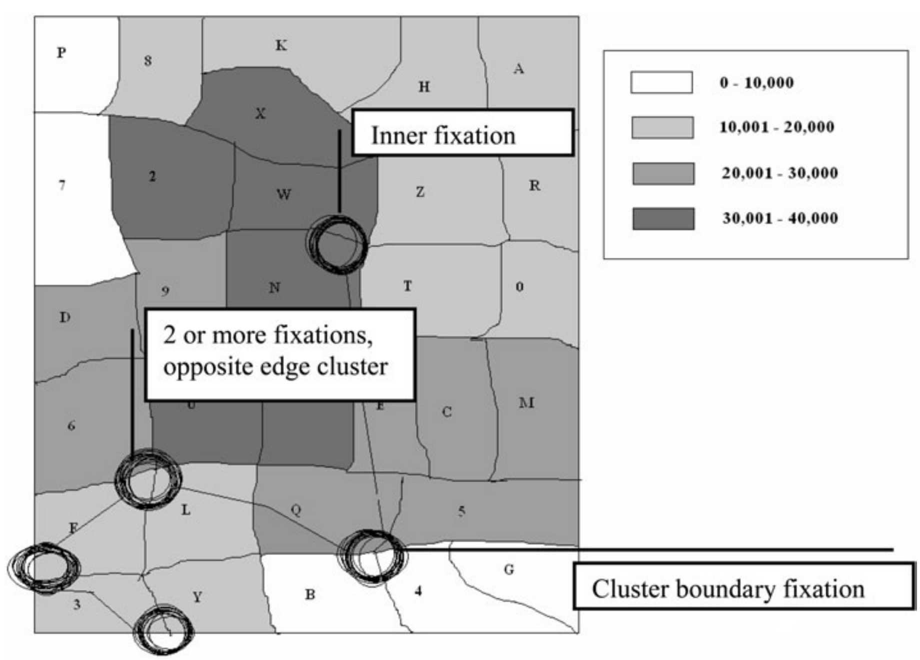

Figure 4. Examples of coded fixations in Experiments 2 and 3; the different shades of gray were actually different colors. The actual figure can be viewed at http://dx.doi.org/10.1037/1076-898X.14.1.36.supp.

counties qualified as a visual cluster (a group of same-colored counties in the graph is referred to as a *cluster*; the actual formation of the cluster with explicit fixations is referred to as a *visual cluster*). Second, if a consecutive sequence of fixations was to at least two edges of opposite boundaries of a single cluster plus the center of the cluster, this was coded as a visual cluster. A second independent coder coded 25% of all the eye track data. Interrater reliability was calculated using Cohen's kappa, $\kappa = .965$, $p < .001$, with interrater agreement at 97.3%.

Performance measures. Participant responses were graded to measure performance. For specific extraction questions, the responses were graded as correct or incorrect depending on whether they identified the specific value associated with the target county in question. For integration questions, responses were graded on a 1–5 scale depending on how well the information in the graph was synthesized. A score of 1 was assigned to responses that simply identified areas in the graph, whereas a score of 5 was assigned to responses that identified several different areas in the graph and synthesized this information to form a coherent representation. The details of this grading system can be found in Appendix B.

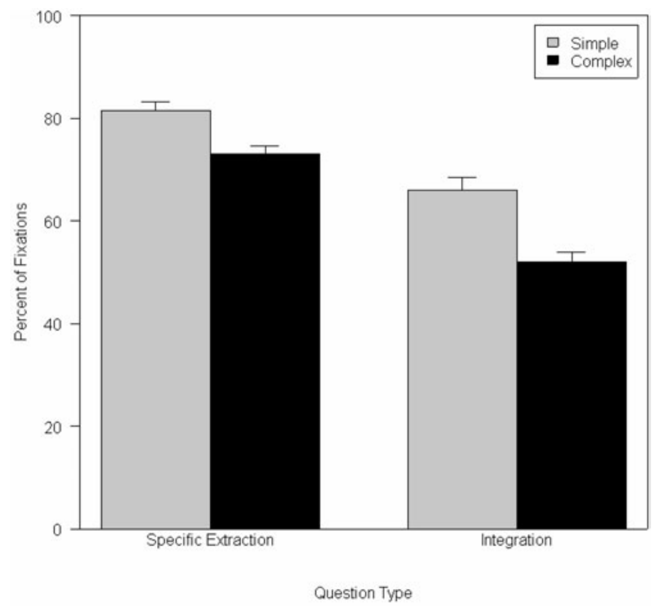

Figure 5. Inner fixations by question type and complexity.

Results and Discussion

The coded eye fixation data were analyzed in relation to the total number of fixations to the graph; thus a percent of fixations relative to the location of all fixations to the graph was calculated. Performance on the specific questions was very accurate (greater than 98%) with no difference between simple and complex graphs ($p > .9$); this performance measure will not be discussed further. Performance on the integration questions is discussed later in this section.

Location of fixations and pattern recognition processes. We first examined the number of inner fixations by graph complexity and question type. The main effect of question type was significant: Specific questions elicited more inner fixations as compared to integration questions, $F(1, 16) = 65.6$, $MSE = 87$, $p < .001$, $\eta^2 = .8$. The main effect of complexity was significant as well: The complex graphs elicited fewer inner fixations as compared to simple graphs, $F(1, 16) = 81.5$, $MSE = 26$, $p < .001$, $\eta^2 = .84$. The interaction of question type and complexity was significant, $F(1, 16) = 10.3$, $MSE = 12.3$, $p < .01$, $\eta^2 = .39$. The interaction was driven by the few inner fixations in the complex graph when integration questions were answered, as shown in Figure 5.

Next, we examined the number of cluster-boundary fixations by graph complexity and question type. The main effect of question type was significant: Integration questions elicited more cluster-boundary fixations than specific questions, $F(1, 16) = 74.5$, $MSE = 52.7$, $p < .001$, $\eta^2 = .82$, as Figure 6 shows. The main effect of complexity was significant: The complex graphs elicited more cluster-boundary fixations than the simple graphs, $F(1, 16) = 82.5$, $MSE = 25.9$, $p < .001$, $\eta^2 = .84$. The interaction was significant, $F(1, 16) = 19.2$, $MSE = 14.3$, $p < .001$, $\eta^2 = .55$. The interaction was driven by the large number of cluster-boundary fixations in the complex graphs when integration questions were answered.

Together, these results clearly show that different perceptual processes occurred during the pattern recognition stage when specific and integration questions were answered. Specific questions elicited primarily inner fixations reflecting the process of searching for the specific county of interest. Integration questions elicited more cluster-boundary fixations and fewer inner fixations as compared to the specific questions. The increased number of cluster-boundary fixations supports the visual integration hypothesis; graph readers looked to the cluster boundaries to form visual clusters. One thing to note is there were still a large number of inner fixations when integrating information. This issue was further explored in Experiment 3.

Evidence of forming visual clusters. Next, we examined whether the cluster- boundary fixations were in service of forming visual clusters. There was no evidence of visual cluster formation when specific extraction questions were answered. When answering integration questions, graph readers formed significantly more clusters in the complex graphs ($M = 4.8$, $SD = .2$) than the simple graphs ($M = 2.1$, $SD = .1$), $F(1, 16) = 61.33$, $MSE = 1.07$, $p < .001$, $\eta^2 = .78$. These results show that, at the perceptual level, visual clusters were being formed when integration questions (but not extraction questions) were answered. When graph readers integrate information, the pattern of visual cluster formation mirrors the cluster-boundary fixation data as well. More complex graphs elicited more cluster-boundary fixations and more visual clusters. Next, we attempt to relate visual cluster formation to cognitive processes.

Connecting visual integration and cognitive integration. To tie visual integration during the pattern recognition stage to cognitive integration, we sought to examine the relationship between the eye movement data and what participants actually said (the verbal data). First, we compared the number of visual clusters formed at the perceptual level to the number of qualitative extractions at the verbal level; a paired sample t test demonstrated that they were not significantly different, $t(16) = .68$, $p = .51$. Further, we ran a standard multiple regression predicting the number of qualitative extractions at the verbal level from cluster- boundary and inner fixations. The overall regression equation was significant, $F(2, 31) = 38.62$, $p < .001$, and accounted for 71% of the

variance in qualitative areas discussed at the verbal level; cluster-boundary fixations loaded significantly, $t(31) = 4.47$, $p < .001$, whereas inner fixations did not load significantly, $t(31) = 1.14$, $p = .26$. The regression equation is as follows:

$$\text{Qualitative verbal extractions} = 1.34 + (.12 \times \text{cluster boundary fixations}) + \varepsilon$$

An additional regression analysis was conducted with the inner fixations that were not directly on county labels (i.e., nonreading fixations) to determine whether these fixations played a role in predicting qualitative verbal extractions. In this regression, qualitative extractions were being predicted from cluster-boundary and nonreading fixations. The results of the regression equation did not change; cluster-boundary fixations were the only significant predictor. These analyses are important for two reasons. First, they show that cluster- boundary fixations, not inner fixations, are critical to integration. Second, they show that these visual integration processes are tightly linked to cognitive integration. These analyses support the hypothesis that visual clusters formed during pattern recognition are then used to reason about the graph using cognitive integration. Next, we look for explicit evidence of cognitive integration and the direct comparison of visual clusters to each other.

Perceptual transitions and cognitive integration. To determine how graph readers performed integration at the cognitive level, we examined the transitional patterns in the eye movement data. We focused on the action immediately following the formation of a visual cluster. If cognitive integration was used, graph readers should have transitioned from one cluster to another as opposed to just relating the visual cluster to the referent (the legend). The percentage of cluster-to-cluster and cluster-to-legend transitions are displayed in Figure 7. The main effect of complexity was not significant, $F(1, 16) = 1.17$, $MSE = 86.31$, $p = .3$, $\eta^2 = .13$. The main effect of transition type was significant; there

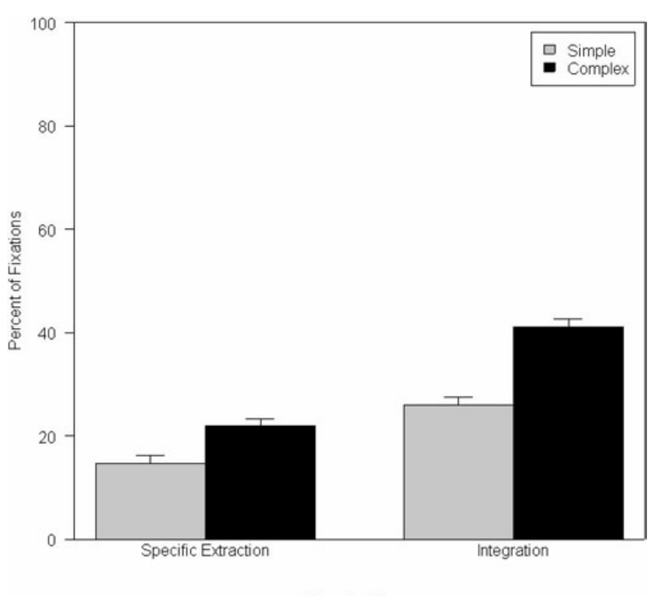

Figure 6. Cluster boundary fixations by question type and complexity.

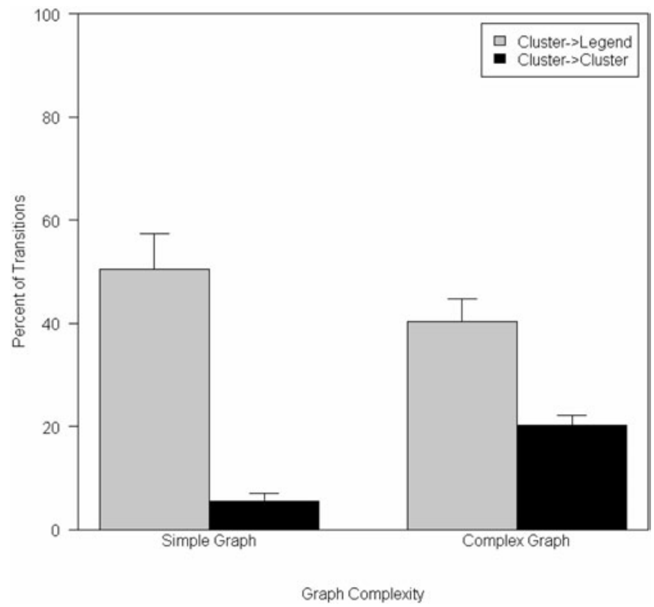

Figure 7. Perceptual transitions for integration questions.

were more cluster-to-legend transitions as compared to cluster-to-cluster transitions, $F(1, 16) = 36.17$, $MSE = 498.31$, $p < .001$, $\eta^2 = .63$. The interaction was significant, $F(1, 16) = 16.41$, $MSE = 160.37$, $p < .001$, $\eta^2 = .45$. Tukey's HSD post hoc comparisons reveal this interaction was driven by no statistical difference in cluster-to-legend transitions between simple ($M = 50.41$, $SD = 15.4$) and complex graphs ($M = 40.41$ $SD = 10.2$); however, there were significantly more cluster-to-cluster transitions in the complex graphs ($M = 20.29$, $SD = 4.8$) as compared to the simple graphs ($M = 5.41$, $SD = 1.1$), $p < .001$.

Even though there were more cluster-to-legend transitions than cluster-to- cluster transitions overall, there were still a substantial number of cluster-to-cluster transitions in the complex graphs (approximately 20%). Thus, there is some support for cognitive integration. At one level, the small number of cluster-to-cluster transitions in the simple graphs is not very surprising. Because there were only three actual clusters in the simple graph, there are likely to be a small number of cluster-to-cluster transitions.

Cognitive integration and quality of response. To determine whether cognitive integration was related to how well graph readers were able to answer the integration questions, we examined the correlations between the perceptual transitions and the quality of answer ratings. There was no difference in the quality of response for simple ($M = 3.51$, $SD = .98$) and complex graphs ($M = 3.53$, $SD = .94$), $F(1, 16) = .03$, $MSE = .15$, $p = .9$, $\eta^2 = .002$, thus, we collapsed across complexity. Overall, cluster-to-cluster transitions significantly correlated to the quality of answer ratings ($N = 17$), $r = .55$, $p < .05$. The cluster-to-legend transitions did not significantly correlate ($N = 17$), $r = -.12$, $p = .6$. The significant correlation between the cluster- to-cluster transitions and the quality of answer suggests that cognitive integration is an important component of the integration process. Explicit visual cluster comparisons seem to be important to integration, whereas cluster-to-legend transitions do not seem to be as relevant.

Summary. We first focused on differences in pattern recognition processes based on question type. Specific extraction questions elicited inner fixations reflecting the search process, while integration elicited cluster-boundary fixations suggesting visual cluster formation. Focusing on the integration process, the cluster-boundary fixations were shown to be in service of explicit visual cluster formation. Further, these perceptual processes (specifically the cluster-boundary fixations) predicted the number of qualitative extractions at the verbal level connecting visual and cognitive integration. Finally, the perceptual transition analysis showed that cluster-to-cluster transitions were part of integration and significantly correlated to the quality of answer.

With the combination of these analyses, we illustrated the integrative process of explicitly forming visual clusters, using visual integration, and then reasoning with these clusters by directly comparing them, using cognitive integration. There are, however, a few issues that are unclear. First, the number of inner fixations observed when answering integration questions was relatively high; second, we expected more evidence of cognitive integration. One possible reason for both of these concerns is that graph readers could have been attracted to reading the county names on the graphs (MacLeod, 1991; Stroop, 1935) despite the fact that this information was not used to answer the integration questions. This would account for the large number of inner fixations. The transition analysis, which hinged on the very next action after forming a cluster, may have been affected as well; a distraction caused by reading county labels would greatly impact this analysis. In Experiment 3, we removed the county labels to further examine the visual and cognitive integration processes.

Experiment 3

Experiment 3 focused solely on integration questions; the main purpose was to remove county names from the choropleth graphs (a) to determine whether inner fixations were a necessary part of visual integration and the formation of visual clusters and (b) to examine whether we could find stronger evidence of the cognitive integration process. Removing county labels also allowed us to create even more complex choropleth graphs (i.e., greater number of actual clusters in the graph). Thus, we could examine how visual and cognitive integration were influenced by this greater complexity as well.

Integration questions may elicit some inner fixations in service of referencing and determining the size of visual clusters. However, we believe that the large number of inner fixations in Experiment 2 may have been artificially high due to reading and that these fixations were not a necessary part of integration. Based on the literature on object segmentation, graph readers should be fixating on cluster boundaries (Bravo & Farid, 2002; Schyns & Oliva, 1994). Thus, in Experiment 3, we do not expect many inner fixations.

Method

Participants. Sixteen George Mason undergraduate psychology students (nine women and seven men) participated in Experiment 3 for course credit.

Materials. Each of the graphs displayed the population of 70 counties; the counties did not have any labels to distinguish between counties (see Figure 8). The size of the counties differed,

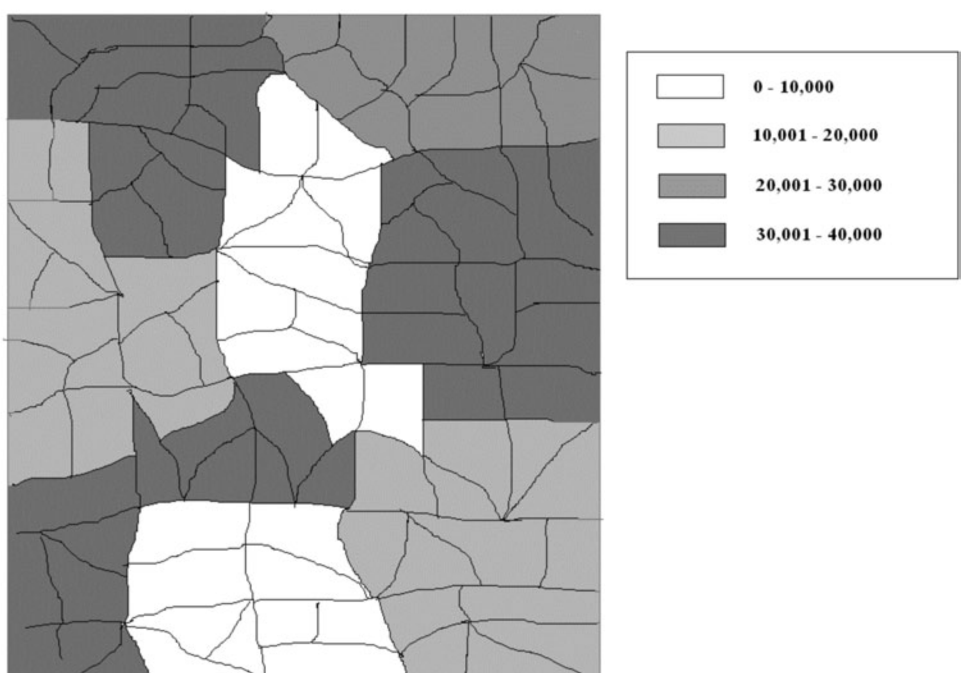

Figure 8. Example of a graph used in Experiment 3; the different shades of gray were actually different colors. The figure can be viewed at http://dx.doi.org/10.1037/1076-898X.14.1.36.supp.

with the smallest county subtending .85° of visual angle and the largest subtending 1.2°. The sizes of the graphs were the same as in Experiment 2. A total of 20 different graphs were generated: 10 graphs had four clusters (simple graphs) and 10 had eight clusters (complex graphs). The graphs were displayed in the center of the computer screen, and the same eye tracker used in Experiment 2 was used in Experiment 3.

Design. Graph complexity was examined in a within-participants design. Participants answered an integration question for the 10 simple and 10 complex graphs.

Procedure. The procedure was the same as in Experiment 2.

Coding scheme. The eye data were coded just as they were in Experiment 2. A second independent coder coded 25% of the eye data. Interrater reliability was calculated using Cohen's kappa, $\kappa = .958$, $p < .001$, with interrater agreement at 97.6%.

Performance measures. The performance measures were the same as in Experiment 2.

Results and Discussion

The eye fixation data were analyzed in relation to the total number of fixations to the graph; a percent of fixations (inner or cluster boundary) relative to the location of all other fixations to the graph was calculated.

Location of fixations. There were more cluster boundary fixations than inner fixations, $F(1, 15) = 38.57$, $MSE = 27.64$, $p < .001$, $\eta^2 = .72$. The complex graphs elicited more cluster-boundary fixations than simple graphs, $F(1, 15) = 33.71$, $MSE = 22.30$, $p < .001$, $\eta^2 = .69$. The interaction of fixation location and complexity was significant, $F(1, 15) = 34.082$, $MSE = 7.41$, $p < .001$, $\eta^2 = .69$, as Figure 9 shows. This was driven by the large number of cluster-boundary fixations in the complex graphs.

Removing the county names from the graphs had an effect on the way graph readers integrated information. These results show that graph readers made many more cluster-boundary fixations than inner fixations when they integrated information; this was particularly true in the complex graphs. Thus, the large number of inner fixations in Experiment 2 was likely due to the existence of county names (presumably a Stroop-like effect) and may not have been necessary for information integration.

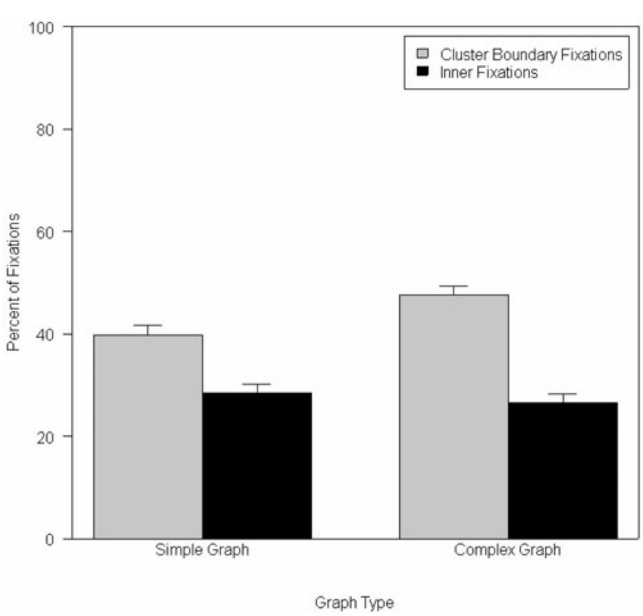

Figure 9. Location of fixations to the graph.

Connecting visual integration and cognitive integration. We sought to tie the visual integration and cognitive integration processes by using eye movement data to predict the quality of the verbal responses. The regression equation predicting qualitative responses from inner and cluster-boundary fixations formulated in Experiment 2 was applied to the eye movement data in Experiment 3. The predicted number of qualitative extractions derived from this equation was correlated to the actual number of qualitative extractions; the predicted and actual qualitative extractions correlated at $r = .71$, $p < .01$. The cluster-boundary fixations accounted for 50% of the variance in qualitative extractions discussed at the verbal level.

This cross-validation makes two important points. First, it stresses the importance of cluster-boundary fixations to the visual and cognitive integration processes. The reduced number of inner fixations in Experiment 3 did not influence the visual integration processes. Second, it illustrates the robustness of the visual clusters. Even though the test set was based on less complex graphs that contained county labels, the equation accounted for a large percent of the variance in qualitative extractions in Experiment 3. This suggests that the visual and cognitive integration processes established in Experiment 2 have scaled up for the more complex graphs. Next, we examined cognitive integration by focusing on perceptual transitions.

Perceptual transition analysis. The percent of visual cluster-to-visual cluster transitions and visual cluster-to-legend transitions were examined. The main effect of complexity was not significant, $F(1, 15) = 1.65$, $MSE = 19.21$, $p = .22$, $\eta^2 = .1$. The main effect of transition type was significant, $F(1, 15) = 6.47$, $MSE = 675.82$, $p < .05$, $\eta^2 = .3$; graph readers made significantly more cluster-to-legend transitions than cluster-to-cluster transitions. As Figure 10 shows, the interaction was significant, $F(1, 15) = 62.04$, $MSE = 53.06$, $p < .001$, $\eta^2 = .81$. Tukey's HSD post hoc comparisons reveal that this interaction was driven by significantly more visual cluster-to-visual cluster transitions in the complex graphs ($M = 35.88$, $SD = 7.8$) as compared to the simple graphs ($M = 22.94$, $SD = 5.2$), $p < .01$, and fewer visual cluster-to-legend transitions in the complex graphs ($M = 38.06$, $SD = 7.9$) than the simple graphs ($M = 53.81$, $SD = 10.4$), $p < .01$. There was no statistical difference between the number of visual cluster-to-visual cluster transitions and visual cluster-to-legend transitions in the complex graphs.

The complex graphs elicited more cluster-to-cluster transitions than the simple graphs, suggesting that as graphs became more complex more synthesis using cognitive integration was required. Thus, in the complex graphs, graph readers formed more visual clusters and made more comparisons among these clusters to integrate information. Next, we examined whether the quality of response was related to the cognitive integration process.

Performance and cognitive integration. There was no difference in the quality of answer ratings between simple ($M = 3.8$, $SD = .8$) and complex graphs ($M = 3.9$, $SD = .9$), $F(1, 15) = .34$, $MSE = .08$, $p = .6$, $\eta^2 = .02$; thus, we collapsed across complexity. The number of cluster-to-cluster transitions significantly cor-

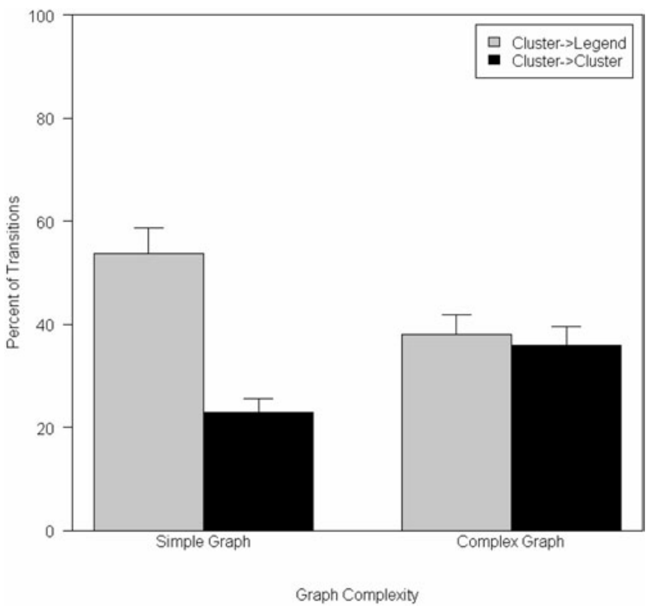

Figure 10. Perceptual transitions.

related to the performance ratings ($N = 16$), $r = .56$, $p < .05$. The number of cluster-to-legend transitions did not correlate to the performance ratings ($N = 16$), $r = -.43$, $p = .1$. This finding replicates the findings from Experiment 2 and shows that the quality of response is closely tied to the amount of cognitive integration. When graph readers made more explicit comparisons, the quality of answer was more integrative.

Summary. The location of fixations analysis and the regression equation analysis suggest inner fixations were not a necessary part of integration; it is the cluster-boundary fixations that are important. The perceptual transition analysis showed strong evidence of cognitive integration, especially in the complex graphs. Finally, the correlation between visual cluster-to-visual cluster transitions and graph performance suggests that visual clusters are a core component of early graph comprehension. These analyses together clearly illustrate the integration process: Graph readers formed visual clusters by examining cluster boundaries and then compared these visual clusters to formulate a coherent representation of the graph.

General Discussion

The goal of this article was to closely examine the perceptual and conceptual processes underlying specific information extraction and integration. For specific information extraction, our experiments confirmed the task analytic theories' account. For integration questions, our theoretical framework introduced two components that are needed to form a coherent representation of the graph: visual and cognitive integration. Verbal protocols and eye movement data provided strong support for both of these components. Visual integration involved the explicit formation of visual clusters of information. Cognitive integration involved the explicit comparison of these visual clusters to the referents and critically to other visual clusters to form a coherent representation. Thus, the visual clusters formed during visual integration served as objectlike units that could then be used to reason about the graph during cognitive integration.

Further, as graph complexity increased, more visual clusters were formed and explicitly compared to synthesize the information in the graph.

Theoretical Implications

Previous theories of integration have suggested two major processes that are different from the processes used to extract specific information. Carpenter and Shah (1998) suggested that multiple cycles of processing are required to integrate information, and Gillian and Lewis (1994) and Wickens and Carswell (1995) stressed the importance of spatial processes. The visual and cognitive integration components that have been highlighted in this article can be unified with these other theories.

The focus of the Carpenter and Shah (1998) framework was on illustrating multiple cycles of processing through pattern recognition and interpretation stages. Our results support the multiple cycles as evidenced by the verbal protocols from Experiment 1. Further, the visual integration processes that were explicitly examined in this article are a detailed description of the processes that occur during pattern recognition. The cognitive integration process of directly comparing visual clusters is a novel process that can be added to the interpretation stages of the Carpenter and Shah framework. In addition to relating visual clusters to the referents, we have shown that visual clusters are explicitly related to each other within the cycles of processing. The process of explicitly comparing visual clusters is a critical component of integration.

Gillian and Lewis' (1994) model and the proximity compatibility principle (Wickens & Carswell, 1995) stress the importance of a spatial component for integration. Our framework suggests that spatial processing is required for both visual and cognitive integration. Specifically, during visual integration graph readers must spatially cluster data points to form coherent objects, and during cognitive integration these clusters must be spatially compared to each other. Further, the spatial demands increase with complexity of the graph.

Together, the previous theories and our framework provide a general process model for integration in graphs. This general process model can be applied to graph types other than choropleth graphs; however, the specific perceptual processes underlying these general mechanisms are likely to be different. The general process model contains three steps. First, during a pattern recognition stage, visual integration occurs: Groups or clusters of information are explicitly formed. Second, during the interpretive stage, cognitive integration occurs: The visual clusters formed during visual integration are related to their referents and are compared to each other to form a coherent representation. Finally, this process is cyclical, and the number of iterations is heavily dependent on the complexity of the graph. More complex graphs will require more cycles of visual and cognitive integration; thus, there is more interleaving between the visual and cognitive components.

Guidelines for Facilitating Integration in Graphs

Several articles provide recommendations for improving graph design (Bertin, 1983; Carpenter & Shah, 1998; Gillian, Wickens, Hollands, & Carswell, 1998; Kosslyn, 1989; Shah & Carpenter,

1995; Shah & Hoeffner, 1992; Shah, Mayer, & Hegarty, 1999; Pinker, 1990). Our framework can also be used to provide specific guidelines for improving integration processes in graphs; specifically, we focus on recommendations for complex visualizations (primarily color-coded graphs) where several data points need to be represented. The overarching goal in designing graphs for efficient integration should be to reduce the number of cycles of processing required to interpret the graph.

To facilitate visual integration, graphs should be designed such that visual clusters of information can be easily formulated. Based on our framework, we suggest three ways of doing this. First, the boundaries of clusters of data should be highlighted such that they are easily identifiable. Fixating on these boundaries to form objects (which are later used in cognitive integration) is a critical component of visual integration, and these boundaries should be highly salient to facilitate this process. A straightforward way of doing this is bolding the boundaries. Second, the color schemes used to code the data should allow easily distinguishable visual clusters. Perceptually linear palettes (Spence, Kutlesa, & Rose, 1999) where color is varied in a single dimension (e.g., varying shades of gray) should not be used to code data, because they can make unique visual clusters harder to identify. Rather, spectral color palettes (e.g., rainbow colors) that allow for easy differentiation between colors should be used. Finally, unnecessary labels or markings should be removed, because they can impact the process of forming visual clusters. These labels can increase the number of fixations required to form explicit clusters and thus increase the amount of time and processing required.

To facilitate cognitive integration, graphs should be designed to reduce the amount of processing needed to reason with the visual clusters formed during visual integration. There are at least two straightforward ways to do this. First, the association between the color-coded data and the legend should be intuitive to reduce the number of cluster-to-legend transitions. Empirically, the experiments presented here show that graph readers spend a fair amount of time looking to the legend after forming visual clusters. If graph readers can intuitively associate a quantitative value with a visual cluster without having to look to the legend, this can reduce processing requirements. Second, the number of uniquely coded variables in the graphical display should be reduced (e.g., do not have too many color codes). Our framework suggests that when there are more clusters in the graph the cycles of processing scale with complexity; more cluster-to-cluster transitions and cluster-to-legend transitions are required. Thus, if the number of coded variables can be reduced by having more general categories of data, graph readers can form a select number of visual clusters and synthesize this information more easily.

References

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

Bravo, M. J., & Farid, H. (2002). Object segmentation by top down processes. *Visual Cognition, 10,* 471–491.

Brewer, C., & Pickle, L. (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers, 4,* 662–681.

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

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic task model of graphical perception. *Human Factors, 34,* 535–554.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Assocation, 79,* 531–554.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* (2nd ed.). Cambridge, MA: MIT Press.

Gillian, D. J. (1995). Visual arithmetic, computational graphics, and spatial metaphor. *Human Factors, 37,* 766–780.

Gillian, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors*, 36, 419–440.

Gillian, D. J., Wickens, D. D., Hollands, J. G., & Carswell, C. M. (1998). Guidelines for presenting quantitative data in Human Factors and Ergonomics Society publications. *Human Factors, 40,* 28–41.

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

Herrmann, D. J., & Pickle, L. (1996). A cognitive subtask model of statistical map reading. *Visual Cognition, 3,* 165–190.

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

Lewandowsky, S., & Behrens, J. T. (1999). Statistical graphs and maps. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 513–549). Chichester, UK: Wiley.

Lewandowsky, 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.

Liu, Y., & Wickens, C. D. (1992). Use of computer graphics and cluster analysis in aiding relational judgment. *Human Factors, 34,* 165–178.

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative approach. *Psychological Bulletin, 109,* 163–203.

Peebles, D., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors, 45,* 28–46.

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

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 124,* 43–61.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Education Psychology Review, 14,* 47–69.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91,* 690–702.

Simkin, D., & Hastie, R. (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association, 82,* 454–465.

Spence, I., Kutlesa, N., & Rose, D. L. (1999). Using color to code quantity in spatial displays. *Journal of Experimental Psychology: Applied, 5,* 393–412.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 12,* 643–662.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies, 53,* 827–850.

Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M.

Hegarty, B. Meyer, & H. Narayanan (Eds.), *Diagramatic representation and inference* (pp. 206–220). Berlin: Springer-Verlag.

Trickett, S. B., & Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, & N. Swaboda (Eds.), *Diagrammatic representation and inference* (pp. 286–300). Berlin: Springer-Verlag.

Trickett, S. B., & Trafton, J. G. (2007). A primer on verbal protocol analysis. In D. Schmorrow, J. Cohn, & D. Nicholson (Eds.), *Handbook of virtual environment training.* Westport, CT: Praeger Security International.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37,* 473–494.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Appendix A

Coding Scheme Used in Experiment 1

Quantitative Extractions

Any specific quantitative information extracted from the graphs was coded as a quantitative extraction. This could be a specific numerical value or a range of values. Examples:

> The population of Victorville County is 30,457.
>
> The population of Janis County ranges from 21,549 to 37,457.
>
> The populations of the counties in the northwest corner are 52,337.

Qualitative Extractions

The extraction of general conceptual information from the graph was coded as a qualitative extraction. Common instances of graph readers extracting qualitative information was when graph readers assign low, medium, or high descriptors to a certain area of the graph or when they simply refer to the color of a general area of the graph without relating the color to the specific legend value. Examples:

> The cluster of counties in the center has a medium population value.
>
> There is a large area of blue and a small area of orange.
>
> Victorville County has a medium low population range.

Search

The process of search was coded for if the graph reader made explicit references to the search process. There must be clear evidence that the graph reader is actually searching. Thus, there must explicit statements like "I don't see it" or "I am looking for. . ." Examples:

> Victorville, Victorville, Victorville, I don't see Victorville.
>
> Let's see. I am looking for Janis County. Where is Janis County?

Reasoning

Any inference that was made that went beyond the basic data that was displayed in the graph was coded as reasoning. The data presented in the graph were relatively context free, thus references to city or country areas in relation to population values are making inferences about the graph. Reasoning involves making statements that clearly go beyond the basic concepts that are represented in the graph. Examples:

> Since the outside seems to be the country area, the center will grow.
>
> The most populated area is probably the city center, and the surrounding less populated areas are likely to be the suburbs that will grow.

Cognitive Integration

Explicit comparisons or relationships formed with the information that was presented in the graph was coded as cognitive integration. These comparisons must be explicit; thus the statement there is a largely populated area in the west and an area of medium population in the north is not cognitive integration. These are two qualitative extractions. To be coded as cognitive integration, the information must be linked by conceptual relation. Examples:

> The highly populated area in the center is much larger than the highly populated area in the upper left.
>
> The cluster of counties on the left are nearly double the size of the counties in the lower right corner.

Appendix B

Performance Rating System for Integration Questions in Experiments 2 and 3

Integration questions were rated on a 1–5 scale based on how well the information in the graph was synthesized. Descriptions of the grading system with examples from the actual protocols given by participants are listed.

Score = 1

A few regions of the graph were identified with no synthesis at all. Example:

> In this map the most populated area is in the southwest corner; the middle section is the least populated.

Score = 2

A few regions of the graph were identified and some statement about the relative size of the regions was made or there was some comparison process. Example:

> The largest population is in the northwest corner and it occupies about a quarter of the graph; there is a small group of counties in the south that are the least populated.

Score = 3

Several regions of the graph were identified with clear evidence of some synthesis among the areas identified including some reference to relative sizes of the areas. Example:

> The center of the graph is the least populated, but the size of this region is substantially larger than the most densely populated area in the northwest.

Score = 4

Several regions of the graph were identified and synthesized with explicit comparisons and relations. The relative sizes of most of these regions were identified as well. Example:

> In the northernmost area is a large group with the lowest population, but going south a little bit is the densest area, which is small. Just below that is a medium-sized group that is the secondmost dense and this extends west and almost covers the length of the map.

Score = 5

There was overwhelming evidence that the graph reader had formed a coherent representation of the graph. Several areas of the graph were identified and the synthesis of this information included direct comparisons and information about the sizes of the regions identified. Example:

> OK, the small group in the northwest is a very low population, but next to it is a very densely populated group that is twice the size. To the east of that is another group of the lowest population; this is a very small group. Directly below this, adjacent to the border, is a. . . the large group in the map that is orange, which is the middle population density range. The large area extends almost half the length of the map and occupies the most space.

Received March 23, 2007
Revision received December 4, 2007
Accepted December 17, 2007 ■