Graph-based Reasoning: From Task Analysis to Cognitive Explanation

David Peebles (djp@psychology.nottingham.ac.uk)
Peter C.-H. Cheng (pcc@psychology.nottingham.ac.uk)
ESRC Centre for Research in Development, Instruction and Training,
Department of Psychology, University of Nottingham, Nottingham, NG7 2RD, U.K.

Abstract

Models of graph-based reasoning have typically accounted for the variation in problem solving performance with different graph types in terms of a task analysis of the problem relative to the particular visual properties of each graph type (e.g. Lohse, 1993; Peebles, Cheng & Shadbolt 1999, submitted). This approach has been used to explain response time and accuracy differences in experimental situations where data are averaged over experimental conditions. A recent experiment is reported in which participants' eye movements were recorded while they were solving various problems with different graph types. The eye movement data revealed fine grained scanning and fixation patterns that are not predicted by standard task analytic models. From these eye-movement studies it is argued that there is a missing level of detail in current task analytic models of graph-based reasoning.

Introduction

The ability to retrieve and reason about information in graphs and diagrams is a skill which requires the complex interaction of three primary elements: the cognitive abilities of the user, the graphical properties of the external representation, and the requirements of the task. Several frameworks have been proposed to understand interactive behaviour of this sort. In the area of graph-based reasoning, Peebles, Cheng & Shadbolt (1999, submitted) have proposed the GBR model incorporating these three factors. Gray (2000; Gray & Altmann, 2000) has proposed the *Cognition-Task-Artifact triad* within which to characterise interactive behaviour in the related context of human-computer interaction. This latter framework has recently been further developed by Byrne (in press) to encompass the perceptual and motor capabilities of the user, termed *Embodied Cognition*.

The main aim of these models and frameworks is to aid the development of detailed cognitive models of the cognitive, perceptual and motor processes involved in the tasks under study. Constructing cognitive process models that are grounded in cognitive theory allows the incorporation and testing of relevant cognitive factors such as the required declarative and procedural knowledge, the strategies adopted, and the limitations of working memory. This approach contrasts with that of *cognitive task analysis* which simply specifies the cognitive steps required to perform the task.

In the area of graph-based reasoning, Lohse (1993) developed the GOMS class of task analysis techniques (Card, Moran, & Newell, 1983; Olson & Olson, 1990; John & Kieras, 1994) by including additional cognitive parameters to produce a cognitive model which simulates how people answer certain questions using line graphs, bar graphs and tables. Lohse's model was based on the assumption that graph knowledge is represented as graph schemas (Pinker, 1990) which allow the recognition and interpretation of different classes of graph. Included in a graph schema are task-specific rules that define sequences of procedures for retrieving information from the graph given a particular information-retrieval task. Lohse's model predicted the time to answer a given question by assuming that people scanned the graphical representation in a manner which produced an optimal sequence of eye movements that minimized the number of saccades and fixations to reach the target location.

In the *Graph Based Reasoning* (GBR) model (Peebles et al., 1999, submitted), a similar set of assumptions was employed to explain several results of experiments investigating the factors affecting reasoning with *informationally equivalent* (Larkin & Simon, 1987) graphs of different types from the same general class; Cartesian coordinate (x–y) graphs. Figure 1 shows the types of graph used in our experiments. The graphs are informationally equivalent as the both encode the same two functions between time and the variables A and B. The *Function* graph in Figure 1a represents time on the *x* axis and the A and B variables on the *y* axis whereas the *Parametric* graph in Figure 1b represents the A and B variables on the *x* and *y* axes respectively while time is plotted as a parameterizing variable along the curve.

Although the two graphs assign different variables to their axes, they would be considered similar in several important ways identified in the literature. Firstly, both are *Cartesian* graphs using a two dimensional coordinate system to relate quantities and represent magnitudes. It is likely, therefore, that both graphs invoke similar general schemas and interpretive processes (Pinker, 1990; Kosslyn, 1989). Secondly, both are simple line graphs and consequently share many of the same general interpretive rules. Furthermore, it is likely that inferences from both graphs are influenced by the same set of biases (Carpenter & Shah, 1998; Gattis & Holyoak, 1996; Shah & Carpenter, 1995). Finally, the graphs are informationally equivalent as they have been generated from the same data set.

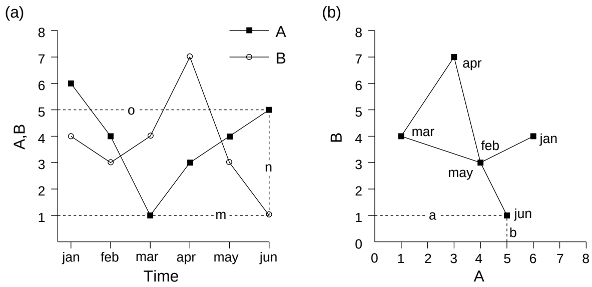

Figure 1: Informationally equivalent function and parametric graphs

Despite these similarities, however, in previous experiments we have demonstrated that for a wide range of questions, parametric and function graph users differ substantially in both the time it takes to respond and in their rates and patterns of errors (Peebles et al., 1999, submitted). The GBR model has been successful in explaining why such differences occur with these graph types despite their many common properties. Using the graphs in Figure 1 as an example, we found in our experiments that when participants were asked to retrieve the value of A when the value of B is 1, responses from parametric graph users were significantly more rapid and accurate than those from the function graph users. The GBR model explains these differences in terms of the optimal visual scan path the users follow through the graph. The variability in responses is apparent from the sequence of hypothesised saccades in the two graphs. In Figure 1a, the sequence of saccades is *m, n, o*, whereas in Figure 1b the process requires just two saccades, as shown by the line sequence *a, b*. The higher probability of an erroneous response using the function graph was explained by the additional number of possible incorrect saccades that the function graph users may make.

Although these optimality assumptions are useful in that they provide an account of differences in mean RT and error data for the different graph conditions, it remains an open question, however, whether they gloss over important cognitive and strategic factors at an individual level. For example, graph users may be required to re-encode items of information that have been lost from working memory during the course of processing. In addition, given that graph users are aware that information is available for re-scanning at all times, it is possible that they may make a strategic decision to trade off additional saccades for a reduction in working memory load. If this is the case, then the current analyses may miss out an important level of detail which sheds light on the cognitive load that these tasks are imposing and the strategies by which graph users optimise their retrieval procedures. Furthermore, information at this level of detail will provide valuable constraints on cognitive models of these reasoning processes.

To address these issues, we devised an experiment in which participants were asked to solve some simple tasks using different graph types of the same general class which, based on the optimality assumptions above, would be predicted to produce different response patterns. These predictions can be elaborated in terms of an optimal sequence of fixations required to solve the given task. To test these optimality assumptions and predictions, therefore, some of the participants' eye movements would be recorded as they solved the problems.

One of the most common tasks carried out when using a graph is to elicit the value of one variable corresponding to a given value of another. This task was chosen for the experiment as it is so widely performed and because the procedures involved are relatively simple. The knowledge required to carry out these tasks is primarily the sequence of fixations required to reach the *given location* in the graph representing the given value of the given variable and then from there to the *target location* representing the corresponding value of the required variable. In previous research, however, we have discovered that the effectiveness of a particular graphical representation for retrieving the required information depends on the details of the task, i.e. which variable is given and which is sought (Peebles et al., 1999, submitted).

Experiment

Method

Participants and materials Forty-four undergraduate and postgraduate psychology students from the University of Nottingham were paid £3 to take part in the experiment. The experiment was carried out using two PC computers with 17 in displays. A further four participants from the same population were paid £5 to participate in the eye-movement study. The eye tracker employed in the experiment was an SMI iView system using a RED II desktop pupil/corneal reflectance tracker with a sampling rate of 50 Hz. This system records eye movements at 20 ms intervals remotely from a position in front of the experimental computer display. Although the system contains an automatic head movement compensation mechanism, to further reduce recording error due to head movement, participant's heads were restrained in a frame fixed to the table.

The stimuli used in the experiment were four graphs, shown in Figure 2, depicting the amount (in millions of units) of UK offshore oil and gas production between two decades, 1970–1979 and 1980–1989. The graphs and data sets were designed so that the independent variable (IV—year) and the two dependent variables (DVs—oil and gas) all had ten values ranging from 0 to 9 and that the full range of these values was represented by the data points for oil and gas in both decades.

Participants were seated approximately 80 cm from the 72 ppi computer display. The graphs were 15.5 cm square (including axis labels), corresponding to approximately 11.1° of visual angle. The characters representing variable values were 0.4 cm high (approximately .21° of

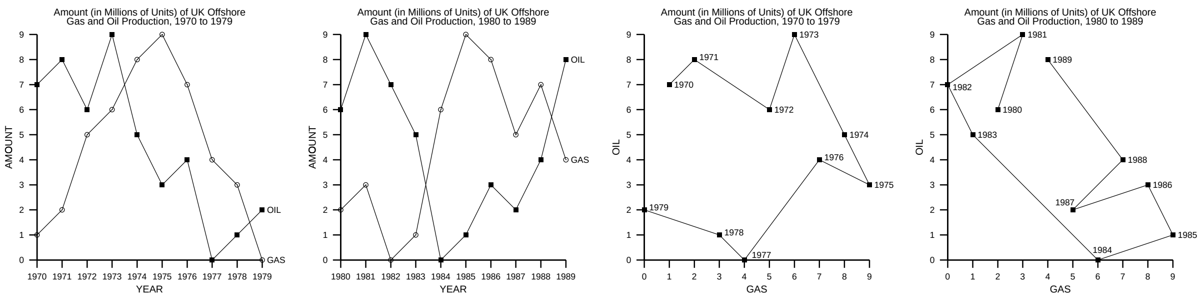

Figure 2: Function and Parametric Graphs Used in the Experiment

visual angle) while those for the axis labels and questions were 0.4 cm and 0.5 cm high (approximately .29° and .36° of visual angle) respectively. Axis ticks were spaced 1.5 cm (approximately 1.1° of visual angle) apart.

The full range of values for each of the variables was used to produce 120 questions. These questions all had the same basic structure and were of three types; DV–DV and DV–IV questions gave the value of one of the dependent variables and required the corresponding value of the second DV or the IV respectively, while IV–DV questions gave a value of the independent variable and required the corresponding DV value to be produced. There were 20 of each of question type and participants were required to answer all 60 for both decade graphs, producing a total of 120 questions.

Design and Procedure The experiment was a mixed design with one between-subjects variable, (graph type) and two within-subjects variables (question type and graph number). Participants were randomly allocated to one of the two graph type conditions producing a total of 22 participants per condition in the main experiment and two participants per condition in the eye movement study. During the experiment, the two graphs were presented alternately with the first graph being selected at random. On each trial, a graph would be presented with a question above it. The questions were presented in a form so that the minimum amount of text was shown. For example, the question GAS = 2, OIL = ? requires the value of oil when gas is equal to 2 to be found. When a year value was required, the final items of text in the question would be YEAR = 197? or YEAR = 198? depending on the current graph being presented and participants were instructed beforehand to enter only the final number of the target year. Each element of the question was centered on a co-ordinate point which remained invariant throughout the experiment with approximately 3.5 cm (approximately 2.5° of visual angle) between the centres of adjacent text items. Together with the graph and question, a button labelled *Answer* appeared in the top right corner of the window. Participants were instructed to click on this answer button as soon as they had obtained the answer to the question. Response times were recorded from the onset of a question to the mouse click on the answer button. When this button was clicked upon, the button, graph and question were removed from the screen and a circle of buttons labelled clockwise from 0 to 9 appeared centered on the answer button. Participants entered their answers by clicking the appropriate number button. When the number button was clicked, the next graph, question, and answer button appeared on the screen. This method was devised so that participants in the eye movement study would not have to take their eyes away from the screen to enter answers, as would be the case if using the keyboard.

Before starting the experiment, participants were given as much time as necessary to become familiar with the two graphs in their condition and were also provided with an opportunity to practice entering numbers using the circle of number buttons and the mouse. Participants were asked to answer the questions as rapidly and as accurately as possible

Results

Response accuracy and latency data The proportions of correct responses and mean response times (RTs) for each of the question types for the two graphs in each condition are presented in Figure 3. Confirming the relative simplicity of the experimental tasks, the data reveal high levels of accuracy for all three question types in both graph conditions. An ANOVA on the response accuracy data, however, revealed a significant effect of question type $F(2, 239) = 28.187$, $p < 0.01$, $MSE = 0.123$ indicating that some types of question were generally more demanding than others. The nature of this effect can be clearly seen in Figure 3. In both graph conditions, more errors were made carrying out the DV–DV task than the other two while the IV–DV task was the most accurately responded to.

While there is little variability in the accuracy of responses between conditions, the time taken by participants in the two groups to make these responses varies significantly both between conditions and within each condition according to the type of question being attempted. An ANOVA on the RT data revealed significant effects of question type $F(2, 239) = 18.447$, $p < 0.01$, $MSE = 4974038$, and graph number $F(1, 239) = 5.76$, $p < 0.05$, $MSE = 1223302$ and significant interactions be-

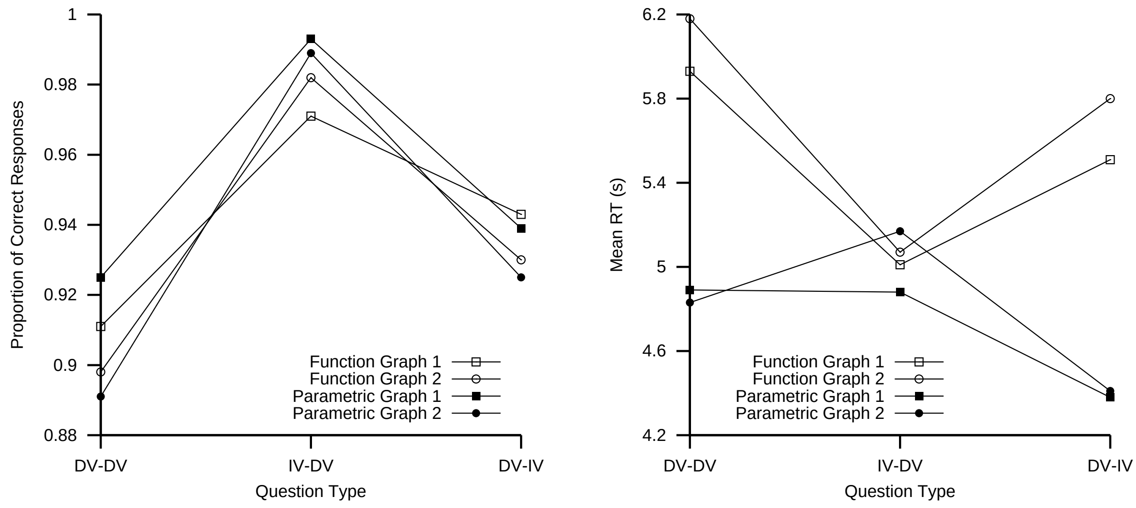

Figure 3: Plots of mean correct responses and RTs for function and parametric graph conditions for each question type

tween graph type and question type $F(2, 239) = 36.314$, $p < 0.01$, $MSE = 9791754$ and between graph type, question type and graph number $F(2, 239) = 3.913$, $p < 0.05$, $MSE = 466423$. The nature of these effects and complex interactions is apparent in Figure 3. In both conditions, it takes approximately 5 s to read the question and retrieve the required DV value for a given year. However, to carry out the reverse task and find the year corresponding to a given DV value takes, on average, over 1 s longer when using the function graph than when using the parametric graph. A similar disparity in RT is found when the task is to retrieve a DV value corresponding to a given DV value.

In both conditions, errors are evenly distributed over experiment trials. The mean proportion of correct responses over the first 10 trials for function and parametric graphs is .91 and .94 respectively. Over the course of the experiment, the mean RT for both conditions reduced by approximately 2 s, the rates of these reductions being described by power functions with similar slopes.

To analyse the results of the experiment, the display was divided into five regions in a manner similar to that employed by Carpenter and Shah (1998). The regions, shown in Figure 4, were the same for all four graphs and define the relevant units of the display for the fixation analysis: *question, graph pattern, x-axis, y-axis,* and *answer* buttons.

The pattern of RT data from the experiment can be explained by the GBR model using the optimality assumptions and fixation predictions outlined above. The significant increase in time to answer DV–IV questions using the function graphs is due to the fact that in the parametric graphs, the target values are positioned next to the given location so that the additional cognitive and perceptual processes required to fixate on the target location are not required. In this case the optimal sequence of fixations is predicted to be: *question, axis, graph, answer* whereas that for the function graphs is: *question, axis, graph, axis, answer*.

The DV–DV questions are of the same type as the example question given in the introduction and so the smaller mean RT in the parametric condition can be accounted for in terms of the previous explanation, namely, that to reach the target location in the function graphs requires an additional saccade and fixation and the associated cognitive operation to retrieve a further step in the process. So, the optimal sequence of fixations for parametric graphs is predicted to be: *question, axis, graph, axis, answer*, whereas that for the function graphs is: *question, axis, graph, graph, axis, answer*.

For the IV–DV questions, the relative rapidity with which function graph users are able to answer these questions compared to others is due to the fact that they are able to rapidly identify the given year on the x axis and then carry out the two step process of identifying the target point on the correct line and retrieving its value from the y axis. The optimal sequence of fixations for this procedure is: *question, axis, graph, axis, answer*. The data show that this procedure takes approximately the same time as the corresponding procedure for the parametric graphs which requires the search of the given year in the graph and the retrieval of its value from the target axis, the optimal fixation sequence of this procedure being: *question, graph, axis, answer*.

The results of the main experiment show that, despite the numerous similarities that exist between function and parametric graphs, the type of graph used can significantly affect the time it takes to retrieve the required information and that this effect is dependent on the nature of task. The experiment also showed that the probability of retrieving incorrect information depends on specific details of the task, i.e. which variable is given in the question and which variable value is being sought. The GBR model explains these differences in terms of a detailed task analysis and the assumption of an optimal scan path through the graph to the target location.

Eye movement data To analyse the eye movement data, the raw *x* and *y* co-ordinate data from the eye tracker were aggregated into *gazes*—sequences of consecutive fixations on a display region unbroken by fixations in other regions (Carpenter and Shah, 1998). The minimum duration of a gaze was defined as 100 ms as this value was sufficiently large to eliminate most saccades, short fixations and noise in the data while still capturing all the relevant fixations. The data from each participant were analysed so that gazes of 100 ms or more in each region were recorded and a scan path consisting of the sequence of gazes for each question was produced.

Several interesting patterns emerge from the analysis of these gaze sequences. Firstly, the average number of transitions between regions for all questions types, shown in Table 1, is consistently greater than the optimal number predicted by the GBR model. For all of the question types, and irrespective of the type of graph being used, participants made, on average, between three and four additional transitions in order to reach the solution. In the majority of cases, these additional transitions were between the axes and the graph and the question and the graph as participants rarely fixated upon the answer region until entering an answer. In 31% of all trials, participants made at least one additional gaze on an axis after having previously fixated upon that axis and then the graph. A detailed visual analysis of the raw eye movement data for these trials revealed that in most cases, participants had fixated upon a given axis value and then proceeded to the plot point in the graph corresponding to that value. Upon reaching this point, an additional saccade was then made to the axis to check that the value was in line with the point.

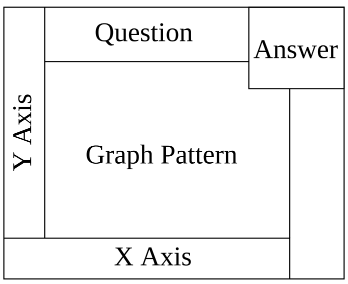

Figure 4: Five regions of the display defined for the fixation analysis

From the eye movement data analysis, it is clear that, although the participants did, in general, solve the various problems by following the optimal gaze paths characterised by the GBR model, they made considerably more gazes than is predicted by the model. Although it is likely that many of these additional transitions are due to checking procedures of the sort outlined above, it is possible that common patterns in the gaze sequences indicate limitations of working memory or problem solving strategies adopted by graph users. For example, in 62.7% of all trials and irrespective of the question type being attempted, participants made at least one additional gaze on the question after having initially gazed upon the question and subsequently the graph. This pattern suggests two possible explanations. The first is that participants have initially encoded the three elements of the question but are required to re-encode certain parts of it that are unable to be retrieved from working memory due to the cognitive load involved in carrying out the problem solving procedures. The second explanation is that participants have adopted a strategy by which only the initial part of the question is encoded and the second part is encoded only when required. According to this explanation, in the majority of trials, participants effectively break the problem into two sections, the first to get to the given location in the graph, the second to move from the given location to the target location corresponding to the solution. It is also possible that the observed gaze patterns may result from a combination of these factors if, during the course of the experiment, participants adopt the above strategy in order to minimise the number of question element retrieval failures.

Table 1: Mean number of gaze transitions between display regions for Function and Parametric graphs observed (Obs) for each question type, and the optimal (Opt) number predicted by the GBR model

Question	Function		Parametric	
Type	Obs	Opt	Obs	Opt
DV–DV	7.66	5.0	8.21	5.0
IV–DV	7.86	5.0	8.90	4.0
DV–IV	8.05	5.0	8.05	4.0

Discussion

Reasoning with Cartesian graphs involves a complex interaction between the perceptual and cognitive abilities of the reasoner, the visual properties of the graph, and the specific task requirements. Models of graph-based reasoning (e.g. Lohse, 1993; Peebles et al., 1999, submitted) have largely focussed on providing a detailed analysis of the task in relation to the the visual properties of the graph and explaining differences in performance in terms of the interaction of these two elements. These models have been successful in accounting for variations in aggregate RT data between users of different graph types by characterising an optimal sequence of fixations based on the task analysis that will achieve the goal. Error data is also explained by hypothesising sets of plausible deviations from these optimal sequences.

To produce detailed cognitive models of graph use grounded in cognitive theory, however, then the third, cognitive element of the triad must be fully incorporated into these accounts. The explanatory and predictive power of cognitive models in complex interactive domains compared to cognitive task analyses has been demonstrated (e.g. Gray, John, & Atwood, 1993). By incorporating such cognitive factors as the user's knowledge, strategies and working memory capacity into graph-based reasoning models, the explanatory and predictive power of these models can be increased and greater insights into the processes and factors affecting these complex interactions can be obtained.

Although the standard experimental variables of RT and error rates provide some information upon which to formulate and test cognitive hypotheses, much richer data is obtained when eye movements are recorded during the experiment. In such a visual domain as graph-based reasoning, eye movements are an important source of information regarding how people acquire and process graphical information and the strategies they adopt when interpreting and working with graphs. This has been demonstrated by Carpenter and Shah (1998) in their analysis of eye movements in graph comprehension tasks which revealed the cyclic nature of the pattern recognition and cognitive processes involved in graph comprehension.

In contrast, the present experiment provides an example of how eye movement data can be used in the analysis of more goal directed graph-based reasoning tasks in which the aim of the interaction is not to simply understand the graph but to retrieve specific information from it. The results of the main experiment showed that the ability of people to retrieve the same information from computationally inequivalent but visually similar Cartesian graphs can be significantly affected by the type of graph used. A plausible explanation of these differences can be provided by the GBR model in terms of an analysis of the task and an assumption of the optimal scan path through the graph to the target location representing the problem solution. These results support and extend the findings of previous experiments (Peebles et al., 1999, submitted) and provide further evidence that the GBR model can account for data that cannot be explained solely in terms of the visual properties of the graphs.

The actual scan paths revealed by the eye movement study show, however, that these optimality assumptions serve as an approximation that can be applied to data aggregated over experimental conditions but which tend to obscure the detailed sequences of saccades made by individuals. It is clear that further research is required to investigate the cognitive factors underlying these saccade patterns in greater detail. It is also clear, however, that cognitive models of graph-based reasoning must incorporate more sophisticated cognitive mechanisms in order to account for these findings.

Acknowledgements

This research is funded by the UK Economic and Social Research Council through the Centre for Research in Development, Instruction and Training.

References

Byrne, M. D., (in press). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*.

Card, S. K., Moran, T. P., & Newell, A., (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Gattis, M., & Holyoak, K. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 231–239,

Gray, W. D. (2000). The nature and processing of errors in interactive behaviour. *Cognitive Science*, *11*, 205–248.

Gray, W. D., & Altmann, E. M. (2000). Cognitive modeling and human-computer interaction. In W. Karwowski, (Ed.), *International encyclopedia of ergonomics and human factors*. New York: Taylor & Francis, Ltd.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, *8*, 237–309.

John, B. E., & Kieras, D. E., (1994). *The GOMS family of analysis techniques: Tools for design and evaluation*. (Tech. Rep. CMU-HCII-94-106). Pittsburgh, PA: Carnegie Mellon University, Human-Computer Interaction Institute.

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

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–100.

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

Olson, J. R., & Olson, G. M., (1990). The growth of cognitive modeling in human-computer interaction since GOMS. *Human-Computer Interaction*, *5*, 221–265.

Peebles, D., Cheng, P. C.-H., & Shadbolt, N. (1999). Multiple processes in graph-based reasoning. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Peebles, D., Cheng, P. C.-H., & Shadbolt, N. (submitted). A model of graph-based reasoning: Integrating the role of visual features, knowledge and search.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*, 43–62.