Multiple Processes in Graph-based Reasoning

David Peebles (djp@psychology.nottingham.ac.uk)[1]
Peter C-H. Cheng (peter.cheng@nottingham.ac.uk)[2]
Nigel Shadbolt (nigel.shadbolt@nottingham.ac.uk)[3]

Department of Psychology, University of Nottingham, Nottingham, NG7 2RD, U.K.

Abstract

Current models of graph understanding typically address the encoding and interpretive processes involved during the course of comprehension and largely focus on the visual properties of the graph. An experiment comparing reasoning with two types of graph is presented. On the basis and scope of existing models, performance with the two graphs would not be predicted to differ substantially. There are substantial computational differences between the graphs, however. It is suggested, therefore, that an adequate model of graph use must incorporate different combinations of visual properties of the graphs, levels of graph complexity, interpretive schemas and task requirements.

Introduction

Graphs are widely used to represent quantitative information in many spheres of human activity. They are commonly encountered in business and the media and are ubiquitous in science and engineering. The essential property common to all graphs is a mapping between the quantitative information being represented and visual dimensions (such as length, size or colour) of specific graphical elements. This visuo-spatial representation of numerical information allows perceptual inferences to be made rather than often more difficult and time-consuming logical reasoning or numerical computation.

Research into graph comprehension has largely focussed on specifying the visual and structural properties of particular types of graph and on providing analyses of the various tasks involved in graph understanding (Bertin, 1983; Cleveland & McGill, 1985; Kosslyn, 1989; Pinker, 1990). Researchers have also developed accounts of various perceptual and cognitive processes involved in graph interpretation (Carpenter & Shah, 1998; Gattis & Holyoak, 1996; Lohse, 1993; Tabachneck, Leonardo, & Simon, 1994).

Although researchers are producing ever more detailed and sophisticated models of graphical perception, the scope of much of this research remains rather narrow. Each study typically examines the properties of only one particular class of graph (e.g. Gattis & Holyoak, 1996; Kosslyn, 1989). In addition, most of these investigations have concentrated on the processes of graph comprehension (e.g. Carpenter & Shah, 1998; Pinker, 1990; Shah & Carpenter, 1995) while less research has been carried out on the actual *use* of graphs for reasoning and problem solving.

[1]Artificial Intelligence Group

[2]ESRC Centre for Research in Development, Instruction and Training

[3]Artificial Intelligence Group

There are different classes and subtypes of graph, each with characteristic features which may be more or less appropriate for representing certain types of information or facilitating different perceptual and cognitive operations. A case may be made both for conducting comparative analyses of different graph types and for studying graph-based reasoning. Firstly, because different types of graph may be used to represent the same information, two graphs may be created which are *informationally* equivalent without necessarily being *computationally* equivalent (for discussions of these issues see Simon, 1978 and Larkin & Simon, 1987). Comparing the behaviour of subjects carrying out the same task using these two graphs may shed light both on subjects' mental representations of particular graphs and the representational and computational properties of the graphs. In addition, by using a wide range of problems, one can produce a rich and varied set of behavioural data which can reveal complex interactions between the graph user's knowledge, the visual properties of the graph and the different perceptual and cognitive operations afforded by the type or class of graph. Graphs and questions may also be produced specifically to test hypotheses about the effect of particular factors. Results of these tests may then be used to specify the factors and components necessary for an adequate model of graph use. In this paper we present results of an experiment designed to identify some of these factors using the methodology outlined above. In a previous study, the class of line graphs was delineated and analysed in terms of an ontology of graphical components shared by the constituent graph types (Cheng, Cupit & Shadbolt, 1998). In this experiment, the properties of two types of line graph identified by Cheng *et al.* are compared. In the

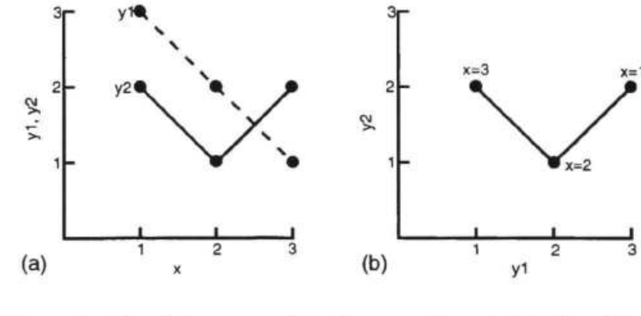

Figure 1: Graph types used in the experiment. (a) *Combined Single Unary* (CSU) graph; (b) *Double Unary Parametric* (DUP) graph

first, called a *Combined Single Unary* (CSU) graph (see Figure 1a), two dependent variables, each of which represents a single unary function of an independent variable plotted on the horizontal x axis, are represented as separate lines superimposed on a single co-ordinate space, the values of which are plotted on the vertical y axis. In the second, named a *Double Unary Parametric* (DUP) graph (Figure 1b), the two dependent variables are represented on the x and y axes and the values of the independent variable are represented by points on the plotted line.

These graphs would be considered similar in various important ways identified in the literature. Firstly, they are informationally equivalent, having been generated from the same data set (Larkin & Simon, 1987). Secondly, they are both diagrams using locational indexing of information (Larkin & Simon, 1987). Thirdly, they are both Cartesian graphs using a two dimensional co-ordinate system to relate quantities and represent magnitudes, so presumably would invoke similar general schemas and interpretive processes (Pinker, 1990; Kosslyn, 1989). Fourthly, they are both simple line graphs so the same set of general heuristics can be used in their interpretation and they would also be affected by the same set of biases (Carpenter & Shah, 1998; Shah & Carpenter, 1995; Gattis & Holyoak, 1996).

As will be demonstrated in the experiment, CSU and DUP graphs differ substantially in computational terms. But it is unclear, given the basis and scope of any one current model of graph comprehension, how the differences in the graphs may be explained. The detailed analysis of subjects' behaviours on selected problems suggests that an adequate model of graph use must integrate different permutations of visual properties of the graphs, levels of graph complexity, knowledge of appropriate interpretive schemas and task requirements.

Experiment

A set of informationally equivalent CSU and DUP graphs was created and a number of questions about the data represented by the graphs were constructed. An additional factor was included in the experiment design—that of graph *complexity*. As a working hypothesis, we have adopted Carpenter and Shah's (1998) equation of the graphical complexity of a graph with the number of plotted lines which can be regarded as distinct functions. Carpenter and Shah have found that the time taken to interpret a graph is strongly related to the complexity of the graph (as they define it). The experiment was also designed, therefore, to determine what effect, if any, increased graphical complexity has on problem solving.

Method

Participants Sixty-four second and third year psychology students from the University of Nottingham were paid to participate in the experiment. Participants were informed that the two people who achieved the most correct answers in the shortest time would also receive an additional payment.

Apparatus and Stimuli The experiment was carried out on an Apple Macintosh computer with a 17 in colour monitor. The primary stimuli used were six line graphs depicting amounts of gold and silver produced by a fictitious mine each month for two consecutive years. Two *simple* graphs of both graph types were created, each presenting the data for one year. A *complex* graph of each type was then created by superimposing the two simple graphs onto a single set of axes. Thus four experiment conditions were produced—simple CSU, simple DUP, complex CSU and complex DUP (henceforth referred to as CSU(s), DUP(s), CSU(c) and DUP(c) respectively). The six graphs are shown in Figure 2.

In order to construct a set of graphs that shared certain visual characteristics to allow them to be used for a single set of problems, the data were generated so that the graphs for each year had a number of similar properties and all graphs satisfied several constraints. For example, each year contained two months with exactly the same amounts of silver and gold production. A set of sixteen questions were produced, eleven of which were seen by all four conditions, the other five being presented after the eleven general questions only to subjects in the complex graph conditions.

Design and Procedure The experiment was a mixed design involving two between-subjects conditions—graph type (CSU or DUP), graph complexity (simple or complex), and one within-subject condition—the year of the graph (1996 or 1997). The 64 participants were randomly allocated to one of the four conditions.

Participants were instructed to answer the questions as accurately and as rapidly as possible and to continuously point to the part of the computer screen at which they were currently looking with a pointer while thinking aloud at all times. Verbal protocols and pointing movements were recorded on audio and video tape respectively. Before starting the experiment, subjects were given a practice trial with six simple questions using graphs of a similar type to their condition in order to familiarise them with the graph type, the equipment and the procedure for answering questions and thinking aloud.

Before answering any questions, participants were shown the graphs and were asked to try to understand what they was about while thinking aloud and pointing to the part of the graph that they were looking at. Subjects in the simple conditions were presented with the graph for each year one after the other in random order. When subjects were familiar with the graphs, they were presented with the questions. In each trial, a question for a particular year together with the graph for that year was presented on the screen. As soon as the subject indicated that s/he had an answer by pressing a key on the keyboard, the graph was removed from the screen and a prompt for the subject to enter a response appeared. Subjects were encouraged to answer all of the questions, entering a "best guess" if they were unsure of the correct answer. The response time (RT) for the initial key press and the subject's response were recorded for each trial. The presentation order of graph year and the questions were randomised. Subjects in the simple conditions saw 22 trials (11 questions for both years) whereas those in the complex conditions received a total of 27 trials (11 questions for both years plus an additional 5 questions).

Results

From a total of 1568 responses, 9 (from different participants) were not accurately recorded. For the purposes of analysis of variance (ANOVA) therefore, each missing response was replaced by a value estimated by computing the cell mean.

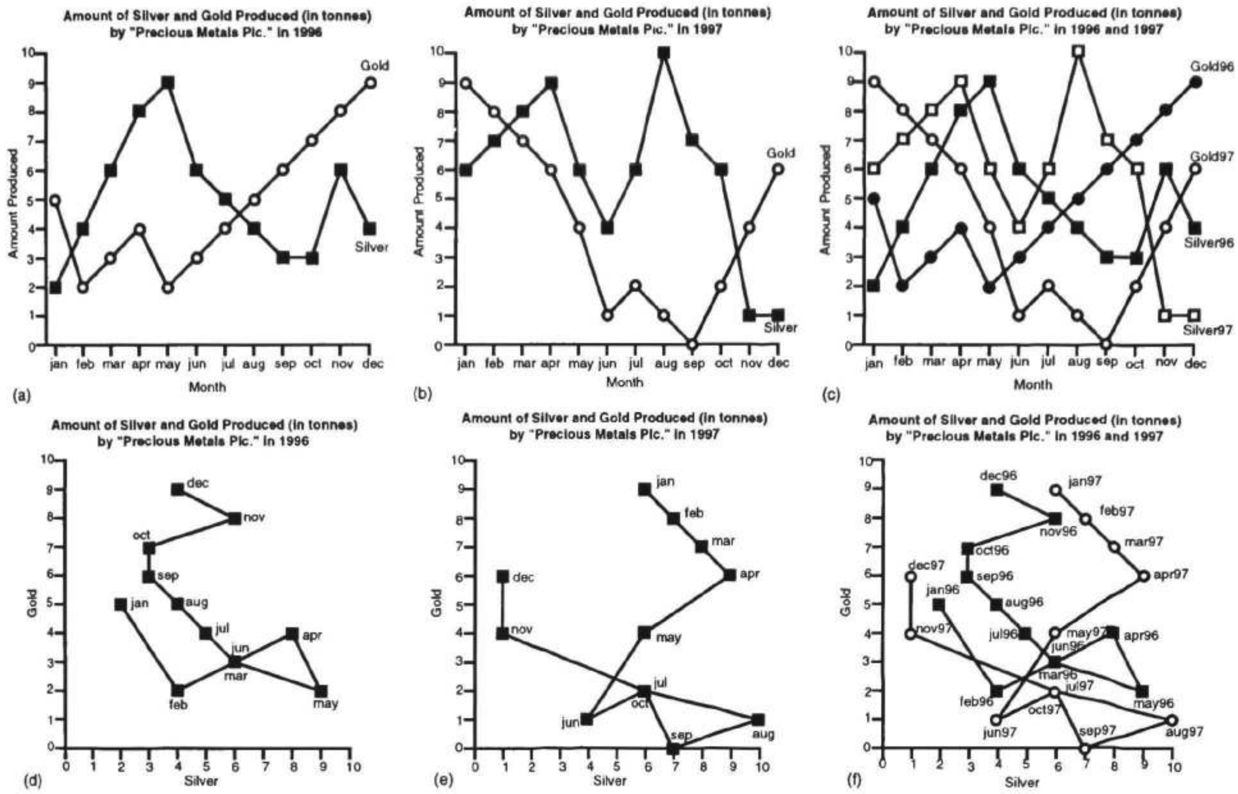

Figure 2: Graphs used in the experiment. (a & b) CSU(s) graphs for 1996 and 1997 respectively, (c) CSU(c) graph for 1996 & 1997, (d & e) DUP(s) graphs for 1996 and 1997 respectively, (f) DUP(c) graph for 1996 & 1997

An ANOVA on the proportions of correct responses for the 11 general questions produced significant main effects of graph type, $F(1,10) = 19.90$, $p < .01$, $MS_e = 0.098$, graph complexity, $F(1,10) = 31.36$, $p < .01$, $MS_e = 0.155$, and question number, $F(10,10) = 74.71$, $p < .01$, $MS_e = 0.368$. Significant interactions were also found between graph type and graph complexity, $F(1,10) = 19.90$, $p < .01$, $MS_e = 0.098$, graph type and question number, $F(10,10) = 20.10$, $p < .01$, $MS_e = 0.099$, and graph complexity and question number, $F(10,10) = 4.01$, $p < .05$, $MS_e = 0.020$. A significant three-way interaction between graph type, graph complexity and question number $F(10,10) = 3.73$, $p < .05$, $MS_e = 0.083$ was also produced. No significant main effect or interaction was found involving the year condition.

An ANOVA on the mean RTs for the 11 general questions produced significant main effects of graph type, $F(1,10) = 10.75$, $p < .01$, $MS_e = 201032528$, graph complexity $F(11,10) = 11.83, p < 0.01, MS_e = 221300928$, and question, $F(10,10) = 121.25$, $p < 0.01$, $MS_e = 2267433216$, and also yielded significant interactions between graph type and graph complexity, $F(1,10) = 18.21$, $p < .01$, $MS_e = 340499488$, and graph type and question, $F(10,10) = 12.65$, $p < .01$, $MS_e = 236482048$. As with the response accuracy data, no significant main effect or interaction involving the year condition was found.

The ANOVAs indicate a complex interaction between the various factors in the experiment. It is clear that both the accuracy of the responses and the time taken to respond is affected by both the type and complexity of the graph being used and the type of problem being solved.

Subjects' graph familiarity After taking part in the experiment, participants were asked to rate the frequency with which they normally encountered information presented in the form just seen by choosing between *High*, *Medium*, or *Low* frequency, and also to rate on a scale of 1 to 9, how familiar they considered themselves to be with the type of graph they had just encountered, where 1 represented *Very Unfamiliar* and 9 indicated *Very Familiar*.

There were considerable differences in the ratings of exposure frequency between the graph types. Only 19.4% of subjects in the CSU conditions rated their exposure frequency to the graph type as low whereas in the DUP conditions, 71.9% of subjects did so.

A similar situation was found in the ratings of graph familiarity. Whereas 74.2% of subjects in the CSU conditions rated their familiarity with the graph type as being 7 or over and none of them rated themselves in the least familiar third of the range, only 28.1% of subjects in the DUP conditions rated their familiarity as being 7 or over while 40.6% of them rated themselves in the least familiar third of the range.

Individual Question Analysis

In this section, we demonstrate the effects of three different factors on problem solving behaviour by analysing the responses to four questions in detail. Each of the questions was selected to highlight the effect of a particular factor.

Question 4: This question illustrates how different graphical representations of the same information can require different procedures to access certain information and how this can result not only in a considerable difference in the distribution of resulting answers but also in a marked difference in the time taken to retrieve the information. Question 4 had the form: *In [1996/1997], when the amount of Gold production is 4, what is the amount of Silver production?*

The graphs were constructed so that two months in each year had amounts of gold production equal to 4 so that at least three answers to the question could be expected—the amount of silver production for either of these months individually or that for both months. Based on a previous analysis of graph-based reasoning protocols, a set of basic perceptual and cognitive operations has been identified which subjects commonly use when solving graph-based problems. Examples of these basic operations include: (a) identifying the axis associated with a given variable, (b) locating a point on an axis corresponding to a given value of the variable, (c) tracing a straight line from a point on an axis until a plotted line is reached, and (d) measuring the distance between two points on a line relative to a given axis. With this set of basic operations, a model of the procedures used to solve specific problems using different graphs can be constructed which can then be used to make predictions about the relative time taken and the probability of different responses being made using a particular graph type. A detailed discussion of this set of basic operations and an outline of such a model will be reported elsewhere. However, to give some idea of the type of analysis which can be carried out and the explanations which can result from this analysis, an outline of two procedures carried out by subjects using the different graph types for 1996 are presented below. Subjects using CSU graphs who give only one answer to this question generally follow this procedure:

1. Identify the axis (a_1) which represents the values of the variable given in the question, (*Amount Produced* axis).
2. Identify the specific point (a_1p_1) on axis (a_1) corresponding to the specific value of the variable given in the question, (*Amount Produced* = 4).
3. Trace a straight line from axis point (a_1p_1) across the graph until the line (l_1) corresponding to the variable and year given in the question is reached at point (l_1p_1), (Point on *Gold* line where *Amount Produced* = 4, [*Month* = 'apr'])
4. Trace a straight line from line point (l_1p_1) until the line (l_2) on the line corresponding to the variable and year required by the question is reached at point (l_2p_1) so that $(l_1p_1 = l_2p_1)$ on the month axis, (Point on *Silver* line where *Month* = 'apr').
5. Trace a straight line from point (l_2p_1) until the axis (a_1) represents the values of the variable required variable in the question is reached at point (a_1p_2), (Point on *Amount Produced* axis from point on *Silver* line where *Month* = 'apr').
6. Read off value of axis location (a_1p_2) and return as answer, (Answer: "Amount of *Silver* produced = 8").

Table 1: Number of responses to each of three answers to Question 4

Answer	Simple CSU	Simple DUP	Complex CSU	Complex DUP
1996				
5 (jul)	1	8	6	9
8 (apr)	9	1	5	3
5 & 8	6	7	4	4
1997				
1 (nov)	1	8	3	11
6 (may)	4	0	8	2
1 & 6	10	8	3	2

In this procedure, as soon as the first point on the plotted line (l_1) is encountered, the corresponding point on the other line (l_2) is identified the value of that point on the (a_1) axis is returned as the answer. To discover the second month with a gold production value of 4, the straight line traced from (a_1p_1) must be extended to reach the second point on the same line. Subjects using DUP graphs who give only one answer to this question generally follow this procedure:

1. Identify the axis (a_1) corresponding to the variable given in the question, (*Gold* axis).
2. Identify the specific point (a_1p_1) on axis (a_1) corresponding to the specific value of the variable given in the question, (*Gold* = 4).
3. Trace a straight line from axis point (a_1p_1) across the graph until the line (l_1) corresponding to the year given in the question is reached at point (l_1p_1), (Point on *Year* line where *Gold* = 4, [*Month* = 'jul']).
4. Trace a straight line from the point (l_1p_1) until the axis (a_2) corresponding to the variable required by the question is reached at point (a_2p_1), (Point on *Silver* axis where *Month* = 'jul').
5. Read off value of axis location (a_2p_1) and return as answer, (Answer: "Amount of *Silver* produced = 5").

Even from this relatively coarse-grained analysis, one can see that the procedure used for the CSU graphs requires an additional step and involves the identification of more lines and points than that used for the DUP graphs. One might expect, therefore, that the time taken by subjects in the CSU conditions to solve this problem will be greater than that of the DUP subjects. More importantly, the graphs in this experiment were constructed so that the first point reached by the procedures (i.e. the month closest to the starting axis) would be different for each condition. Therefore, if the above procedures are followed, the profile of responses given by subjects in the two conditions will differ. The numbers of responses corresponding to the three answers given by subjects in the four conditions are shown in Table 1. The mean RT for the DUP conditions (14.4 s and 16.4 s for the DUP(s) and DUP(c) conditions respectively) was on average 8.04 s faster than that of the CSU conditions (21.8 s and 25.0 s for the CSU(s) and

CSU(c) conditions respectively). In both the simple and complex conditions the differences between the graph types were significant, $t(30) = 2.55, p < .05$ and $t(30) = 2.94, p < .01$ respectively. In the two CSU conditions, 35.9% of the participants gave the answer of two values, 40.6% gave the answer corresponding to the month nearest the starting axis and 17.2% gave the answer corresponding to the alternative month. In the DUP conditions, 32.8% of the subjects gave the answer of two values, 56.3% gave the answer corresponding to the month nearest the starting axis and 9.4% gave the answer corresponding to the alternative month. In both CSU and DUP conditions, therefore, more subjects gave one answer corresponding to the month nearest the starting axis than to the alternative answer or both answers.

This pattern of results, together with an analysis of the video-taped protocols, provides evidence that the procedures outlined above were carried out by the majority of subjects using the two graph types. Question 4 demonstrates that both the speed and nature of a problem solution can be determined by the graphical representation because of the different procedures which users are required to follow in order to access the same information.

Question 5: This question demonstrates how graph-based reasoning processes involving perceptual components can require less time and be less error-prone than those involving numerical computation. Question 5 had the form: *How many months in [1996/1997] is there a difference between Silver and Gold production of exactly 1 tonne?*

Because the two dependent variables are plotted as lines in CSU graphs, differences between the two variables for a given month are represented by a vertical distance between the two lines at a specific value of the x axis. In the CSU conditions, therefore, subjects are simply required to scan the graph to find points on the x axis where the distance between the two plotted lines for a given year are separated by one unit on the y axis. In DUP graphs, however, this spatial representation is not available as the value of each variable for a given month is represented by the co-ordinate location of one particular point. In the DUP conditions, therefore, subjects are required to attend to each point on the plotted line in turn, obtain the gold and silver value for that point from the two axes, and then compute the difference between the two values. One may expect, therefore, that the more elaborate and error-prone procedure used in the DUP conditions would be reflected in lower response accuracy and longer RT scores from subjects in the DUP conditions.

The proportion of correct responses and mean response times for each of the four conditions are shown in Table 2. The average response time for the CSU(s) condition was 17.55 s faster than that of the DUP(s) condition $t(30) = 6.56, p < .001$ while the mean RT of the CSU(c) condition was 15.00 s faster than that of the DUP(c) condition $t(30) = 4.14, p < .001$. The accuracy of the responses from the CSU conditions was also greater than those from the DUP conditions. The average RT for the CSU(c) condition was only 5.41 s slower than that for the CSU(s) condition and the correct response scores for the two conditions was virtually identical.

A study of the video-taped protocols supported the analysis above, revealing that a large proportion of subjects solved the

Table 2: Proportion of correct responses (CR) and mean response time (RT in seconds) for the four conditions to Question 5

	Simple		Complex	
	CSU	DUP	CSU	DUP
CR	.938	.844	.969	.625
RT	18.9	36.5	24.3	39.2

problem using the procedures appropriate for the particular graph representation, whether perceptual or conceptual.

Questions 12 and 13: The purpose of questions 12 and 13 taken together was to investigate the combined effect of two factors—graph complexity and graphical representation—and to determine whether the effects of the former may be mitigated by the latter. The two questions illustrate how the retrieval of certain information (e.g. the maximum and minimum sum, difference and product of the two dependent variables), can be facilitated by a representation in which the information can be found by searching a local region for a specific point, compared to a representation in which the same information can only be retrieved by a process of search and computation.

Question 12 had the form: *In which month in 1996 was the greatest total amount of metal (i.e. both Silver and Gold) produced?* The form of question 13 was: *In which year was the most metal (i.e. both Silver and Gold) produced in any month?*

Questions 12 and 13 were only presented to subjects in the complex conditions. In the DUP(c) condition, the item of information required by both questions (the maximum sum of the two metals) corresponds to a point on a line, the location of which is the furthest to the top right corner of the graph. In the CSU(c) condition, however, the same information is derived by identifying a point on the x axis at which the points on two specific lines have positions on the y axis which result in a combined value greater than any other. Therefore, whereas the correct answer may be found in the DUP(c) condition using a visual search procedure which identifies the location of a single point, the search procedure in the CSU(c) condition requires the additional computation of the sum of the two y axis values.

Question 13 differs from question 12 in that it requires the maximum sum of the two metals to be found across both years. If subjects in the DUP(c) condition identify the correct year by locating a point on one of two lines which is nearest to the top right corner of the graph, they should take approximately the same time to carry out this procedure as for question 12. In the CSU(c) condition, however, the procedure required to answer question 13 is more demanding than that for question 12 because it requires subjects to take all four lines into account as they must find the maximum sum of two metals from two pairs of lines rather than from only one pair in question 12. Therefore, if the procedures outlined above are followed, one may expect not only that subjects in the CSU(c) condition would take longer to produce an answer than those in the DUP(c) condition, but also that a greater discrepancy should be found between the average RTs between the conditions for question 13.

Table 3: Proportion of correct responses (CR) and mean response time (RT in seconds) for the two conditions to Question 12 and 13

	Question 12		Question 13	
	CSU	DUP	CSU	DUP
CR	.938	.875	.800	1.00
RT	36.8	27.0	54.5	25.4

The response accuracy and RT data for questions 12 and 13 are shown in Table 3. Although there was very little difference in the accuracy of the responses between the conditions for question 12, those from the DUP(c) condition were on average 9.86 s faster than those from the CSU(c) condition, $t(30) = 2.33, p < .05$. Mean RTs of subjects in the DUP(c) condition for question 13 were similar to those from question 12. Responses of subjects in the CSU(c) condition, however, were on average 17.62 s slower than those from question 12, $t(15) = 3.63, p < .01$. Subjects in the DUP(c) condition were on average 29.03 s faster and 20% more accurate in their responses to question 13 than those in the CSU(c) condition, $t(30) = 3.40, p < .01$.

Analysis of the video-taped protocols revealed that many subjects in both conditions initialised a search by attending to months at which one of the values was particularly high (May and December). Several subjects in the CSU(c) condition scanned along the x axis from January to December, computing possible candidate solutions until the highest sum was found. Many subjects in the DUP(c) condition, however, rapidly located the correct point without attending to other points on the line, suggesting that they were aware of the general region at which the required information must lie.

Discussion

The results of the experiment show that a number of interrelated factors can significantly affect graph-based reasoning and that these factors cannot be accounted for simply by an analysis of the visual properties of the graph. These results support the claim that an adequate model of graph use must also take into account the specific representational properties of a graph type and computational procedures which are facilitated by particular graphical representations. Question 4 illustrated how different graphical representations of the same data can require different procedures to access the same information and that the time taken to access the information, and even what information is retrieved, can be affected by the procedure followed. Question 5 demonstrated that representing an item of information in a graph as a visual feature such as distance can result in a more rapid and accurate retrieval of that information than when the information must be computed from the visual information. Questions 12 and 13 illustrated three main points. Firstly, both questions showed individually that accessing certain types of information can be facilitated by a graphical representation in which the information is represented as a specific location compared to a representation in which extraction of the information requires computational effort. Secondly, the two graphs taken together showed that graph complexity can have an affect on problem solving performance but that this effect can be mitigated by the type of representation used. Thirdly, Questions 12 and 13 showed that the effect of a particular representation can be large, even when users of the particular graph are relatively unfamiliar with its form.

Acknowledgements

This research is funded by a grant from the UK Engineering and Physical Sciences Research Council and by the UK Economic and Social Research Council through the Centre for Research in Development, Instruction and Training. The authors wish to thank the fourth member of the project, James Cupit, for his comments on earlier drafts of this paper.

References

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps.* Madison, WI: The University of Wisconsin Press.

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

Cheng, P. C-H., Cupit, J., & Shadbolt, N. R. (1998). Knowledge acquisition from graphs: An ontological framework. In B. R. Gaines, & M. Musen (Eds.), *Knowledge Acquisition Workshop, KAW98, 1*, (pp. 1–19). Banff, Alberta: University of Calgary.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analysing scientific data. *Science, 229*, 828–833.

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

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.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle, *Artificial Intelligence and the Future of Testing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Simon, H. A. (1978). On the forms of mental representation. In C. W. Savage (Ed.), *Minnesota Studies in the Philosophy of Science. Vol. IX: Perception and Cognition: Issues in the Foundations of Psychology.* Minneanapolis: University of Minnesota Press.

Tabachneck, H. J. M., Leonardo, A. M., & Simon, H. A. (1988). How does an expert use a graph? A model of visual and verbal inferencing in economics. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 842–846). Hillsdale, NJ: Lawrence Erlbaum Associates.