How People Represent and Reason from Graphs

Lara Webber & Aidan Feeney
Department of Psychology
University of Durham, Queen's Campus,
University Boulevard,
Stockton-on-Tees, TS17 6BH,
United Kingdom

{l.j.webber, aidan.feeney}@durham.ac.uk

Abstract

In this paper we examine how people represent graphical information. We present a constrained graphical reasoning task isomorphic in logical structure to a three-term series reasoning problem. Participants were shown pairs of simple line graphs (premise graphs) and were then required to verify a third line graph (conclusion graph). We found that participants reordered the premise graphs in order to construct integrated representations. The order of the terms in the premises (their figure) modulated the accuracy and speed with which participants subsequently verified conclusions against these representations. These findings suggest a role for analogical representation in graph comprehension and call into question the common assumption that graph comprehension processes may accurately be modelled using propositional representations only.

Introduction

In this paper we will be concerned with how people represent and reason from graphs. Interest in these topics is rising perhaps due to the increasing popularity of graphs as a means of presenting information (for a review, see Zacks et al, 2001). However, the literature on graph comprehension is replete with assumptions that are rarely subjected to empirical test (Cheng, Lowe & Scaife, 2001). Our goal in this paper is to outline, and test, some intuitions about how people represent simple graphs.

In the burgeoning literature on graph comprehension there are a number of common assumptions (see also Scaife & Rogers, 1996). For example, it is commonly assumed that graphical displays can easily convey a lot of information that would otherwise result in large amounts of confusing text (for example, see Hammer, 1995). Another common assumption in the cognitive science literature is that graphical displays are more efficient than sentential representations although this assumption holds, at best, only sometimes (Cheng et al, 2001; Larkin & Simon, 1987).

There are a number of suggestions in the literature as to why graphical displays might be more efficient than sentential representations. According to Larkin & Simon (1987), graphs and diagrams support efficient information search and make explicit information that is left implicit in propositional representations. Of most interest here is Shimojima's (1999) claim that whilst graphical representations are governed by nomic constraints, sentential representations are stipulative in that they are governed by conventions such as grammatical rules. Nomic constraints, according to Shimojima, are natural laws that involve topographical, geometric or physical relations. In other words, diagrammatic representations are in some way analogous to the things that they represent, and the relationships portrayed therein are isomorphic to relationships in the real world. The analogical nature of diagrammatic representations may explain how information that is left implicit in propositional representations is made explicit by a diagram.

If it is the case that, by virtue of their adherence to nomic constraints, diagrammatic representations differ from, and are sometimes more efficient than, sentential representations, then there is a very puzzling paradox in the recent literature on graph comprehension. Several recent attempts to model the processes involved in graph comprehension have assumed that people represent the information conveyed by a graph in propositional form (for example, see Pinker, 1990; Shah & Carpenter, 1995). However, we have just seen that graphs may be more efficient than sentential representations because they are analogous to the things that they represent. The claim that people mentally represent graphical information in a propositional format seems counter-intuitive. The argument that we wish to advance here is that (at least sometimes) people's representation of the information contained in graphs is analogical (see also Feeney et al 2000; Fischer, 2000; Trafton et al, 2000). In what follows we will describe the reasoning and methods that have allowed us to test our claim.

Testing for Analogical Representations

One reason why researchers have assumed that people always represent graphical information propositionally is that it may be convenient, for the purposes of cognitive modelling, to do so. This convenient modelling choice is made more attractive by the dearth of empirical work addressing how people represent graphs. Although it seems strange that the relationship between external and internal representations has received little experimental attention, this gap in the literature may be due to the difficulty of testing the nature of the representations underlying any cognitive task.

In the literature on deductive reasoning (for a review see Evans, Newstead & Byrne, 1993) much debate has focused on the nature of the representations that people construct whilst reasoning. Some theorists claim that people construct analogical representations, such as spatial arrays (Huttenlocher, 1968) or mental models (Johnson-Laird & Byrne, 1991), whilst other posit the use of propositional representations (e.g. Rips, 1994). A key research aim in this literature is the development of research methods designed to reveal the nature of the representations underlying deduction.

One type of deduction that has proved particularly amenable to study is reasoning about relationships. Relational reasoning problems consist of a set of premises that describes the relationships between a set of entities. For example, given the premises in problem 1:

A is taller than B
B is taller than C Problem 1

people might be asked to say which entity is tallest, or to verify a conclusion such as “A is taller than C”. The former task requires people to deduce information from the premises in order to derive a conclusion, whereas in the latter they are required to verify the conclusion. The relational terms used usually determine what conclusion, if any, follows, assuming the premises to be true. The time taken to make these decisions (reaction time - RT) is considered a direct result of inference difficulty and a good indicator of how the problem is represented (e.g. Potts & Scholz, 1975). For Problem 1 above the RT is usually fast and the error rate low.

According to analogical accounts of relational reasoning such as Mental Model theory (Johnson-Laird & Byrne, 1991), when drawing a relational inference, the reasoner constructs an integrated model of the premises that is analogical in structure to the state of affairs described in the premises. For example the premise set described in Problem 1 (which we will refer to as AB: BC) is represented by the model below:

A
B
C

where vertical position represents relative height. As A is taller than B in the world, it occupies a higher position in the mental model representation. This representation enables the reasoner to simply to read off the conclusion AC or A is taller than C. The ease with which this inference may be drawn seems, to us, a good example of the way diagrammatic representations make information explicit (Larkin & Simon, 1987). According to analogical accounts of relational reasoning, problem difficulty is a direct result of how easy or difficult it is to construct an initial representation of the premises.

The evidence for analogical representation in relational reasoning is very strong and the weight of that evidence has recently been added to by fMRI studies that have shown increased activation during relational reasoning of areas in the brain that are associated with spatial and visual processing (Goel & Dolan, 2001; Knauff et al, 2002). One important test of whether an analogical representation underlies performance on a particular reasoning task is the presence of the “Figural Effect”. Johnson-Laird & Bara (1984) define figure as the form the premises take i.e. the arrangement of the end terms (A and C in Problem 1) in relation to the repeated term (B in Problem 1). They found that figure affected performance and that term order modulated the conclusions that people spontaneously produce. The figure AB:BC leads to more conclusions in the direction AC whereas the figure BA:CB leads to more CA conclusions. This means that one figure produces conclusions whose terms appear in the same order as they do in the premises whilst with the other figure, term order in the conclusion is reversed.

According to mental model theory whether or not term reordering occurs depends on how the premises are combined in order to create an integrated analogical representation of the premises. The figure AB:BC is easily combined as the repeated terms occur in sequence, whereas the figure BA:CB requires reordering of the premises to bring the repeated term to the middle before integration. Thus the figure BA:CB would become CB:BA and the conclusion CA would then follow. Johnson-Laird & Bara (1984) found this effect both with syllogisms and three-term series problems. Furthermore this effect has also been found when reasoning with diagrammatic representations in disjunctive reasoning (Bauer & Johnson-Laird, 1993) and in spatial relational inference (Knauff et al 1998).

The figural effect is not predicted by any of the sentential accounts of relational reasoning (Clark, 1969; Hagert, 1984) that have appeared in the literature. Thus, the discovery of the effect in participants’ responses on a task is taken as good evidence that people have represented the information in the premises analogically. We now describe a graph-based reasoning task designed to test for the existence of a figural effect in people’s graphical reasoning.

A Constrained Graphical Reasoning Task

To investigate graphical reasoning we developed a constrained graphical reasoning task in which participants were presented with two simple line graphs simultaneously. Each line graph depicted one of the premises from a 3-term relational reasoning problem. Each of the terms was represented by one of the data points in the line graph. Figure 1 depicts the graphical

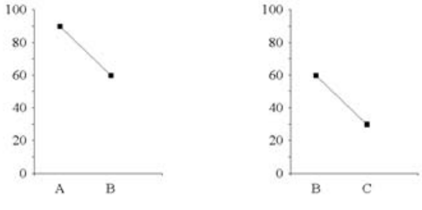

Figure 1: Two simple line graphs with descending slope and end terms separated by the repeated term.

analog of Problem 1.

We manipulated the position of the End Terms so that they were either separate (as in Figure 1) or adjacent (see Figure 2), and the slope of the graphs either ascended left to right (Figure 2) or was descending (Figure 1).

Participants were then shown a third graph specifying a relationship between the end terms (A and C) that they were required to verify. The relationship depicted in the conclusion graph was either valid or invalid, with consistent or inconsistent labelling. Consistent conclusion

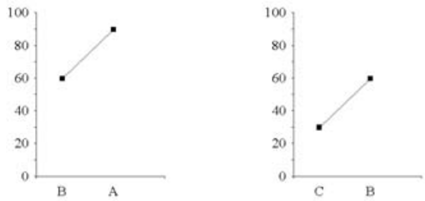

Figure 2: Two simple line graphs with ascending slope and adjacent end terms.

graphs had the same label order as in the premise graphs so that AC was a consistent conclusion for AB:BC whilst CA was inconsistent. Consistency refers to the order of the alphabetic labels only.

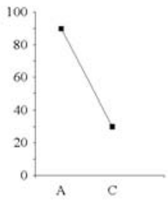

Figure 3: A valid conclusion graph in which the term order is consistent with term order in Figures 1 &2.

If participants represent the graphical information in this task analogically, then they will have to reorder certain premise sets to construct an integrated representation. Thus we should find effects of figure where certain sets of premises lead to more errors and greater RTs than others. For example, if participants construct an integrated model of the premise graphs then the graphs in which the end terms are separated by the repeated term should be easier to represent than the graphs in which the end terms are adjacent. This requirement to swap the terms around in the adjacent trials will incur greater processing costs and thus, is likely to lead to an inaccurate or insufficient representation of the premises for the conclusion validation stage.

In addition to predicting longer inspection times for premise sets in which the end terms are adjacent rather than separated by the repeated term, we also predict effects in the error rates and conclusion verification times. Term order in the conclusion AC is consistent with term order in the premise sets ABBC and BACB. However, in order to construct an integrated model the latter premise set must be transformed into CBBA. Term order in this representation is now consistent with term order in the conclusion CA. Thus, we predict an interaction between the position of the end terms in the premise graphs, and whether the order of the terms in the conclusion is consistent with the order of the terms in the premises. When the end terms are separated in the premise graphs, verification RTs should be shorter and errors fewer for consistent conclusions than for inconsistent conclusions. On the other hand, when the end terms in the premise graphs are adjacent, we predict shorter verification latencies and fewer errors for trials in which term order in the conclusion is inconsistent.

Method

Participants

52 students took part in the experiment of whom 40 were female and 12 were male. Participants' age ranged from 17 to 44 with a mean age of 25.

Materials

The materials consisted of 112 computer-based trials, of which 64 were experimental and 48 were distracter trials. Each trial consisted of two visual displays, a premise display and a conclusion display, which appeared in that order. The premise display (see Figures 1 & 2) contained two simple line graphs, which participants were told always depicted the sales figures (in hundreds of thousands of pounds) for a branch of a company. Each line graph consists of two connected data points with each data point representing a different manager of the branch. The most successful manager was represented by a data point at 900mm, the next most successful by a data point at 600mm and the least successful by a data point at 300mm. Each manager was identified by their initials.

When participants had finished inspecting the premise display they were required to press the space bar on their keyboard. This initiated the display of a second screen containing the conclusion graph (see Figure 3). The conclusion graph specified the relationship between two of the managers. For the experimental trials, this graph consisted of two data points representing the least and most successful managers, which were exactly the same height as in the premise graphs. Reversing the order of the data points and/or reversing the order of the labels of the data points generated four possible conclusion graphs. Of these, two were valid and two invalid, two had labels that were consistent with the order in the premises and two that were inconsistent with that order. There was a total of 4 premise graphs, each with 4 possible conclusions. Four sets of materials were constructed for each combination of premise and conclusion graphs by using different sets of manager initials. This resulted in a total of 64 experimental trials.

The premise graph displays used for the distracter trials were the same as those in the experimental trials. However the conclusion graphs were different. Each conclusion was either a valid or invalid version of one of the graphs in the premise display. Unlike the experimental trials, the labeling remained consistent. There were 16 (8 valid and 8 invalid) conclusion graphs that contained the tallest end term and 16 matched trials containing the shortest end term. To ensure that integrated conclusions did not vastly outnumber non-integrated conclusions, we also included a further 16 trials with conclusions containing short end terms.

Design

We recorded the time taken by participants to inspect the premise graphs as well as the time taken to validate the conclusion graph. We also recorded the number of errors made by each participant. The experiment was entirely within participants and the factors we manipulated with respect to the premise graphs were: Slope (Descending vs. Ascending) and End Terms (Separate vs. Adjacent). With respect to the conclusion graphs we manipulated Validity (Valid vs. Invalid) and Consistency (Consistent vs. Inconsistent).

Procedure

Data was collected from groups of between 3 and 6 participants in separate testing sessions. Each participant sat in front of an IBM-clone computer and monitor with a keyboard. Participants next read a set of instructions for the experiment. Each trial consisted of a fixation cross (duration 1000ms) followed by a premise display that remained on the screen until the participant indicated that they had read it by pressing the space bar. This was then followed by a conclusion display, which remained on the screen until the participant had made a yes/no response using the keyboard. Half of the participants were required to make 'yes' responses with their dominant hand whilst the remainder made 'no' responses with their dominant hand. The between trial interval was 1000ms.

Results

Only the data from the valid trials was included in our analysis. Mean error rates were calculated for each participant across conditions and converted into percentages. Participants who had an overall error rate of the mean error rate (13.28%) for the entire experiment plus one standard deviation (18.76%, total: 32.04%) and over were excluded from all subsequent analysis, (8 participants in total). We used one standard deviation to trim our error data as error rates are measured on a finite scale where 50% indicates chance responding. We trimmed our inspection and reaction time data using two standard deviations as latencies have the potential to be infinite.

Inspection Time Data

Data from trials in which participants made correct responses were analysed using a 2 (Slope) x 2 (End Terms) within participants ANOVA. Any inspection times with a response of below 100ms (there were none) or of more than two standard deviations above the mean (mean: 4779ms, S.D.: 3964, total: 12706ms) were also excluded (3.98% of trials). The mean reaction times for each condition of the experiment are displayed in Table 1. The ANOVA revealed a main effect of End Terms: $F(1, 43) = 8.03$, $MSE = 687507.0$, $p<.007$. Graphs in which the end terms were Adjacent (4344 ms) were inspected for longer than were graphs in which the end terms were Separate (3990ms). This is consistent with our predictions concerning term order.

Table 1. Summary of mean inspection times (ms) for each condition.

	Descending	Ascending
Separate	3911	4068
Adjacent	4306	4382

Reaction time data

Only the reaction times (RTs) for trials in which participants made a correct response were included for analysis. Any trials with a response time of below 100ms (none were recorded) or of more than two standard deviations above the mean (3.9% of trials) were discarded before analysis. The data trimming resulted in three missing values, which were replaced by the mean of the data set (2187ms). A 2x2x2 within participants ANOVA was carried out on the data and the mean reaction times from each cell of the design are displayed in Table 2.

Table 2. Summary of mean reaction times (ms) for each condition.

		Descending	Ascending	
Separate	Consistent	1950	2141	*2045*
	Inconsistent	2201	2333	*2267*
		2057	*2237*	
Adjacent	Consistent	2351	2152	*2252*
	Inconsistent	2197	2171	*2184*
		2274	*2162*	

The ANOVA revealed no main effects. However, the interaction between Consistency and End Terms was significant: $F(1, 43) = 4.71$, $MSE = 391034.2$, $p <.04$. Tests for simple effects showed that when the end terms were separate there was a significant effect of consistency: $F(1, 43) = 11.25$, $MSE = 2162015$, $p <.002$. Consistent conclusions were validated more quickly than inconsistent conclusions. Although the effect of consistency did not attain significance when the location of the end terms was adjacent, examination of the means in Table 2 shows that inconsistent conclusions were validated more quickly than were consistent conclusions. Thus, the interaction is exactly as would be predicted if people construct integrated analogical representations of the information in the premise graphs.

The ANOVA also revealed an interaction between Slope and End Terms: F(1, 43) = 6.29, MSE = 263515.2, p <.02. Tests for simple effects revealed an effect of end terms when the slope was descending: F(1, 43) = 7.63, MSE = 1741018, p<.02. Graphs in which the end terms appeared separate were responded to faster than graphs in which the location of the end terms was adjacent. There was no effect of end terms when the slope of the graphs was ascending: F(1, 43) = .96, MSE = 251223.9, p>.3.

Correct Response Rates

The error data was analysed using a 2(Consistency) x2(Slope) x2(End Terms) within participants ANOVA. The mean error rates for each condition of the design are presented in Table 3.

Table 3: Summary of mean % of errors for each condition

		Descending	Ascending	
Separate	Consistent	2.84	3.98	*3.41*
	Inconsistent	10.22	7.39	*8.81*
Adjacent	Consistent	8.52	6.25	*7.39*
	Inconsistent	2.27	6.25	*4.26*

The analysis revealed no significant main effects. However, as was the case for the reaction time data, the error analysis revealed a significant interaction between Consistency and End Terms: F(1, 43) = 8.49, MSE = 188.276, p <.006. Tests for simple effects show that performance on conclusions that follow on from premise graphs in which the end terms were separated by the middle term is significantly more accurate when the order of the terms in the conclusion is consistent with term-order in the premises: F (1, 43) = 8.65, MSE = 1281960, p <.006. The reverse pattern is observed in the error data for conclusions that follow on from premise graphs in which the end terms are adjacent. That is, trials where term order in the premises and conclusions is inconsistent elicited fewer errors than when the orders were consistent. Although this difference did not attain significance (F(1, 43) = 2.26, MSE = 429.6875, p>.1), the significant interaction between End Terms and Consistency is exactly as would be predicted if participants are constructing analogical representations of the premise graphs.

Discussion

Our findings strongly suggest that people construct analogical representations of the information contained in the premises of our graphical reasoning task. Premise graphs in which the end terms are adjacent take longer to inspect than do premise graphs in which the end terms are separated by the repeated term. Furthermore there is strong evidence of figural bias in the error and RT data. When the end terms in the premises are separate, consistent conclusions are responded to faster and more accurately than inconsistent conclusions. The reverse is true when the end terms in the premises are adjacent. All of these findings are consistent with the idea that people re-order the information in the premises in order to construct an integrated analogical representation. In this representation the relationships amongst items in the world are captured by their relative positions along an axis.

Our results confirm the intuition that, at least sometimes, people's internal representations of graphical information are subject to the same nomic constraints (Shimojima, 1999) as the external graphical representations on which they are based. Thus, we can claim to have demonstrated a correspondence between external and internal representations. This goes some way towards answering the complaint of Scaife & Rogers (1996) that researchers tend to assume a correspondence rather than demonstrating its existence.

Some objections to the generality of our claim are possible. For example, participants in our task inspected the premise graphs with the goal of validating a conclusion concerning the information contained therein. Everyday graph comprehension may not be so goal-directed. Similarly, the graphs we used in our trials were extremely simple. Perhaps people will extract more information and use different representational strategies when the visual display is more complex. Although we plan to address these issues in future work, there is already some relevant work in the literature. For example, Mani & Johnson-Laird (1982), using an incidental-learning paradigm, have found evidence for spatial representation in people's memory for text. In their experiment, people had to decide whether a diagram and a set of verbal premises described the same state of affairs in the world. When the verbal description was determinate (i.e. consistent with only one state of affairs) people's memory for the text was consistent with the idea that they had represented the information using an analogical strategy. This finding suggests that more complex tasks requiring the integration of diagrammatic and verbal information can lead to analogical representations. Mani & Johnson-Laird's results, along with Trafton et als (2000) finding that meteorologists construct qualitative mental models to represent complex weather patterns, increases our confidence that we will find evidence for analogical representation when we ask people to think about richer graphical representations than those we have described here.

We do not wish to suggest that people never represent graphical information in propositional format. Evan a cursory examination of the text-comprehension literature reveals that people represent text propositionally and analogically (for a review see Singer, 1990). Instead, we wish to correct the unfounded assumption that graphs are always represented propositionally. Our data demonstrate that this is not so. Future work should investigate the conditions under which people adopt different representational strategies for working with graphs and diagrams. Our strong intuition is that the goals and abilities of the information processor will determine the nature of the correspondence between internal and external representations.

References

Bauer, M. I. & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, *4*, 372 – 378.

Cheng, P.C-H., Lowe, R. & Scaife, M. (2001). Cognitive science approaches to understanding diagrammatic representations. *Artificial Intelligence Review, 15,* 5-27.

Clark, H. H. (1969). Linguistic Processes in Deductive Reasoning. *Psychological Review, 76,* 387-404

Evans, J. St .B. T., Newstead, S. N. & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.

Feeney, A., Hola, A. K. W., Liversedge, S. P., Findlay, J. M. & Metcalf, R. (2000). How people extract information from graphs: Evidence from a sentence-graph verification paradigm. In M. Anderson, P. Cheng & V. Haarslev (Eds.), *Diagrams 2000, Lecture Notes in Artificial Intelligence 1889,* Berlin: Springer Verlag.

Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs? *Applied Cognitive Psychology, 14,* 151-162.

Goel, V. & Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia, 39,* 901-909.

Hagert, G. (1984). Modeling mental models: experiments in cognitive modelling of spatial reasoning. In T O'Shea (Ed.), *Advances in Artificial Intelligence*. Amsterdam: Elsvier Science Publishers.

Hammer, E. M. (1995). *Logic and visual information*. United States: CSLI Publications.

Huttenlocher, J. (1968). Constructing Spatial Images: A Strategy in Reasoning. *Psychological Review, 75,* 550-560.

Johnson-Laird, P. N. & Bara, B.G. (1984). Syllogistic inference. *Cognition, 16,* 1 - 61.

Johnson-Laird, P. N. & Bryne, R. M. J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates.

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R. & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional MRI study. *Cognitive Brain Research*, 13, 203-212.

Knauff, M., Rauh, R., Schlieder, C. & Strube, G. (1998). Continuity Effect and figural Bias in Spatial Relational Inference. In M.A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah: Lawrence Erlbaum Associates.

Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science*, *11*, 65-99.

Mani, K. & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition, 10,* 181-187.

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

Potts, G. R. & Scholz, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning and Verbal Behaviour, 14,* 439-456.

Rips, L. J. (1994). *The Psychology of Proof.* Cambridge: MIT Press.

Scaife, M. & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45,* 185-231.

Singer, M. (1990). *Psychology of language: An introduction to sentence and discourse processes.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Shimojima, A. (1999). The graphic-linguistic distinction. *Artificial Intelligence Review, 13,* 313-336.

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.

Zacks, J., Levy, E., Tversky, B. & Schiano, D. (2001). Graphs in print. In P. Olivier, M. Anderson & B. Meyer (Eds.), *Diagrammatic representation and reasoning*. London: Springer-Verlag.

Acknowledgements

The authors are very grateful to Bob Metcalf for writing the presentation program described in the paper.