Feeney, A., Hola, A., Liversedge, S., Findlay, J., & Metcalf, R. (2000). How people extract information from graphs: Evidence from a sentence-graph verification paradigm. In P. C. M. Anderson, V. Haarslev (Ed.), *Diagrams 2000, LNAI 1889* (pp. 149-161). Berlin: Springer-Verlag.

How People Extract Information from Graphs: Evidence from a Sentence-Graph Verification Paradigm

Aidan Feeney, Ala K.W. Hola, Simon P. Liversedge, John M. Findlay
and Robert Metcalf[1]

Department of Psychology, University of Durham, Science Laboratories, South Road, Durham
DH1 3LE, United Kingdom

Graph comprehension is constrained by the goals of the cognitive system that processes the graph and by the context in which the graph appears. In this paper we report the results of a study using a sentence-graph verification paradigm. We recorded participants' reaction times to indicate whether the information contained in a simple bar graph matched a written description of the graph. Aside from the consistency of visual and verbal information, we manipulated whether the graph was ascending or descending, the relational term in the verbal description, and the labels of the bars of the graph. Our results showed that the biggest source of variance in people's reaction times is whether the order in which the referents appear in the graph is the same as the order in which they appear in the sentence. The implications of this finding for contemporary theories of graph comprehension are discussed.

1. Introduction

Graphs are a ubiquitous part of everyday life and their use appears to be still on the increase [1]. Despite the proliferation of graphic communication, comparatively little is known about how people extract information from even the simplest of graphs. A central goal of this paper is to examine the processes involved in comprehending simple bar graphs. In general, the attention paid to graphs by psychologists has been disproportionately small relative to the extent of their use. Moreover, much of the work that exists has been concerned either with people's understanding of the graphical representation of complex concepts [e.g. 2, 3] or with the production of a general, high-level, theory of graph comprehension [e.g. 4, 5]. More recently, however, there has been some interest in the detailed cognitive processes underlying the comprehension of very simple graphs [6, 7]. Below we will discuss this work along with an early sentence-picture verification study that we consider to be highly relevant to our experiment.

Our primary interest in this paper is how people form a representation of sentential and graphical descriptions in order to decide whether they agree. This interest stems from the observation that we rarely see, or produce, graphs outside of a linguistic context. That is, most graphs have a title and are accompanied by some text. Furthermore, when a person processes a graph, they very often do this with the objective of verifying that claims made about the graph in the accompanying text are correct. Our starting point, then, is that graph comprehension is not usually an open-ended task. Instead, we view it as goal directed. People comprehend graphs in order to verify claims or to understand some quantitative relationship that is currently the subject of focus.

1.1 The Sentence-Graph Verification Task: A Tool for Studying Goal Directed Comprehension

The approach we adopt to examine goal oriented graphical interpretation is to present participants concurrently with a sentence and graphical display and require them to make a decision as to whether the sentence is an accurate description of the graph. Sentence-picture verification methodology is a paradigm that is highly suited to our needs in this study. In order to perform the task, participants must carry out goal oriented graphical interpretation. Hence, adopting this methodology and carefully manipulating different aspects of both the linguistic statement and the graphical display, will allow us to determine those characteristics of the text and graph that make interpretation more or less difficult.

The most common use of the paradigm has been to investigate how negation is represented and comprehended [e.g. 8, 9, 10]. To our knowledge, the most relevant use of this paradigm has been reported by Clark and Chase [11] who investigated how people verify that a verbal description of the relationship between two referents corresponds to a pictorial representation of the relationship between the same referents. The relationships which Clark and Chase asked their participants to verify were ’above’ and ’below’ whilst both sentences and pictures referred to ’star’ and ’plus’ signs. Participants were shown arrays (see Figure 1) where a verbal description such as *star is above plus* was placed to the left or the right of a simple picture of a star above or below a plus sign and were asked to indicate whether the descriptions matched. Clark and Chase proposed an additive model of the processing stages involved in their sentence-picture verification task that accounted for the influence on people’s verification times of (i) the presence of negation in the description; (ii) the relational term used; (iii) the order of the referents in the sentence; and (iv) whether the sentence and the picture matched. The first assumption underlying their model was that in order to verify a sentence against a picture both must be mentally represented in the same representational format. Secondly, once a representation has been formed of the first description attended to, a representation of the second description will be constructed using the same relational term as was used in the first representation.

Star isn’t above plus	+ *	Array 1
Plus is above star	+ *	Array 2

Figure 1: Examples from picture-sentence verification paradigm [11]

The sentence-graph verification paradigm that we adopted in our experiment was analogous to the sentence-picture verification paradigm used by Clark and Chase and others. To illustrate this paradigm, in Figure 2 we present some sample arrays from our experiment. Each array comprised a sentence specifying a relationship between two referents presented below a bar graph representing a relationship between the same two referents.

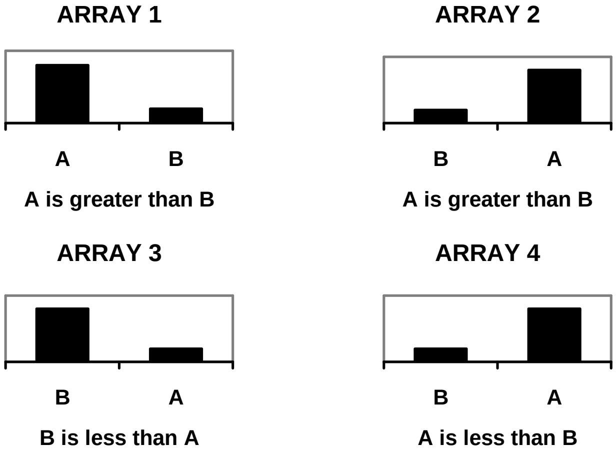


Figure 2: Sample arrays from the sentence-graph comprehension paradigm used in our experiment.

The participant's task was simply to decide whether the statement was an accurate description of the graph. In Figure 2 Arrays 1 and 2 constitute matching trials in which the statement is in agreement with the graph while Arrays 3 and 4 are mismatching.

We manipulated four variables: whether the sentence and graph matched; the order of the bar graph labels (alphabetic or non-alphabetic); the slope of the graph (ascending versus descending); the relational term used in the sentence (*greater*, *less*). These variables produce an exhaustive set of possibilities. In combination they determine whether the order of the referents in the sentence is congruent with the order of the labels in the graph. Whilst the order of the terms agree in Arrays 1 and 3 they are in disagreement in Arrays 2 and 4. We will refer to such arrays as being *aligned* (in the former case) or *non-aligned* (in the latter case).

If verification of sentence-graph displays involves similar cognitive processes to those Clark and Chase suggested are involved in sentence-picture verification, then we would expect differences in reaction time suggesting that people represent the second description they encode using the same relational term as that used in the first description. In other words, if a participant reads the sentence *A is greater than B* they should encode the graphical display in terms of the relational term *greater than*.

Additionally, Clark and Chase argued that picture-sentence verification involves two separate stages of processing: one of encoding and a second distinct stage during which people compare their representations of the sentence and the picture to check for a match. Consequently, if we can extrapolate their approach to sentence-graph verification, we would also expect our reaction time data to provide similar evidence for two distinct stages of processing.

In this section we have considered Clark and Chase's theoretical account of picture-sentence verification and also the implications this theory might have for an account of sentence-graph verification. However, it is important to point out that the processes involved in graph comprehension may be quite different to those involved in picture comprehension. In the next section we will briefly consider some of these differences and then focus on an account of graph comprehension proposed by Pinker.

1.2 Pinker's Account of Graph Comprehension

Even the most simple of bar graphs contains substantially more information than do the pictures illustrated in Figure 1. This extra information comes in the form of various conventional features of the graph. For example, the X and Y axes, the scale on the Y axis, each of the Bar labels and their positions as well as the physical characteristics of the bars of the graph. Given the additional information, we might expect encoding processes and verification processes in our sentence-graph verification task to be more complex than in the simple picture-verification task used by Clark and Chase. In line with this, Pinker [4] and Gattis and Holyoak [6] have claimed that we possess specific schemas for interacting with graphs and that these schemas appear to emerge at a relatively early stage of development [12].

One way to think about how the graph will be processed is provided by Pinker [4] who suggests that it is the interaction between the graph, the task at hand and the information processor's background knowledge that determines the ease with which information may be extracted from a graph. He argues that the processing of graphical information happens in a number of stages. First, visual processes code the graph into a visual array. Secondly, a 'visual description' or propositional representation consisting of predicates and variables is constructed. This description of the observed graph is compared against graph schema stored in memory in order to decide what kind of graph is being viewed. The activated schema aids the extraction of a conceptual message based on the information present in the visual description. At this point in a sentence-graph verification process the participant's representation of the sentence becomes relevant. If the information required to verify the sentence is not present in the conceptual message, the visual description of the graph is interrogated via the activated schema. Sometimes inferential processes may be carried out on the conceptual message itself in order to extract information required to make a verification decision.

In terms of our task, we expect participants to construct a visual representation of the graph. We would also expect a propositional representation to be constructed after which schematic knowledge about types of graphs would be activated automatically. In fact, in our experiment where participants receive many trials consisting of similar graphical representations, schematic knowledge should be activated to a substantial degree. Pinker also argues that graph readers should be able to translate higher-order perceptual patterns, such as a difference in height between a pair of bars, into an entry in the conceptual message extracted from the visual description via the instantiated schema. Accordingly, it should be a relatively simple task for a participant to interrogate the conceptual message with reference to the relationship between the referents described in the sentence.

Whilst Pinker's account goes some way towards outlining the processes involved in graph comprehension, it has been criticised for its generality [13]. It seems to us best thought of as a description of the macro-processes, rather than micro-processes underlying graph comprehension. Consequently Pinker's account is not sufficiently detailed to allow us to make specific predictions concerning our verification paradigm. We anticipate that our sentence-graph verification paradigm will shed light on some of the detailed processes involved in encoding and extracting conceptual information from graphs. For example, it should be informative as to the flexibility of the encoding processes (Is it equally easy to encode different types of graph?). Our paradigm may also provide insight into how graphical information is represented. We anticipate that whilst it is unlikely that there will be comprehension time differences for each of the sentence forms in the study, it is likely that the information contained in certain sentences will be easier to verify against certain graphical representations than against others. Accordingly, the presence of meaningful patterns in our reaction time data should shed light on the processes involved in interrogating and drawing inferences from [4] the visual description of a graph.

2. Experiment

2.1 Method

Participants: 51 undergraduate students at the University of Durham took part in this experiment.

Materials: Our materials were 64 different visual displays such as the one below. Each display consisted of a simple bar graph specifying the position of two referents on a scale and a verbal description of the relationship between these referents. There were sixteen experimental conditions in this experiment and each will be described with respect to the example above. Figure 3 contains a graph that we classified as descending because the slope of its bars descends left to right. To produce ascending graphs we simply swapped the position of the bars. The relational term used in the verbal description in Figure 3 is 'greater than'. Half of the displays used in this experiment employed 'greater than' whilst the other half employed 'less than'. In addition, the order in which the labels appeared in the graph was manipulated. The labels in Figure 3 appear in alphabetic order but for half of our experimental trials this order was reversed. Finally, the descriptions in Figure 3 match. That is, the sentence provides an accurate description of the graphical display. To achieve a mismatch in Figure 1 we reversed the order of the entities in the verbal description.

We constructed four examples of each experimental condition. These examples differed in terms of the labels used (A&B, C&D, L&M, P&Q), the widths of the bars in the graph, the distance between the bars and the difference between the height of the bars. All factors other than the labels were counterbalanced across conditions so

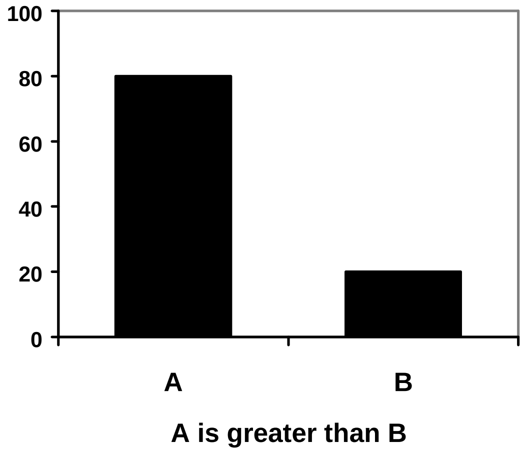


Figure 3: Sample display from our experiment.

we would not expect them to influence our results. Note that the magnitude difference of the bars was always easily discriminable. The order in which the resulting displays were presented was randomised.

Design: The experiment had a 2(Slope) x 2(Label Order) x 2(Relational Term) x 2(Match/Mismatch) within participants design. All participants received 64 trials each requiring them to indicate whether the graph and verbal description of the graph were in agreement.

Procedure: Data was collected from 51 participants in two separate testing sessions (25 participants per session approx.). Each participant was seated in front of a IBM clone computer and monitor and was given a booklet containing instructions and a sample display. Once participants indicated that they understood the instructions they were required to start the experiment. Each trial consisted of an initial display of a fixation cross (duration 1000 ms) followed by the graphical display with the sentence underneath. This display remained on the screen until the participant made a Match/Mismatch decision by pressing one of two buttons. Half the participants were required to make a MATCH response using their dominant hand. The interval between the end of one trial and the onset of the next trial was 1000 ms.

2.2 Results

Reaction Times. For the reaction time (RT) analyses **w**e discarded all trials where participants had made an incorrect response, and all trials with RT's less than 100 ms (none occurred) or greater than two standard deviations from the mean RT for the entire experiment. In total we discarded 8.03% of the data using this procedure, (4.1% errors and 3.9% outliers). We then computed mean RT's for each participant across the remaining trials for each experimental condition.

Table 1: Means RT's (ms) with standard deviations (in italics).

		Ascending		Descending		
Rel-Term	Label Order	Match	Mismatch	Match	Mismatch	
Greater	Alpha	2307	2077	1902	2394	2170
than		*575.8*	*489.5*	*446.9*	*550.1*	
	Non	2319	2194	2043	2379	2234
	Alpha	*585.1*	*470.8*	*560.2*	*550.0*	
		2313	2136	1973	2387	2202
Less	Alpha	2035	2440	2298	2083	2214
than		*555.0*	*526.7*	*579.1*	*600.0*	
	Non	2007	2418	2262	2050	2184
	Alpha	*428.9*	*536.8*	*521.5*	*460.0*	
		2021	2429	2280	2067	2199

We carried out a 2x2x2x2 within participants ANOVA on the data. The means and standard deviations from this analysis are presented in Table 1. The analysis revealed two significant results. First, the main effect of whether the linguistic description agreed with the graph (F(1, 50) = 13.65, MSE = 173580.5, $p < .001$). The mean RT for matching trials was 2146 ms versus a mean of 2254 ms for mismatching trials. Such a difference between matching and mismatching trials has been observed previously in the literature. For example, Clark and Chase [11] observed that for affirmative sentences, RTs for correct responses to matching trials were shorter than RT's for correct responses to mismatching trials.

More novel is our finding of a highly significant interaction between Slope, Relational Term and the Match variable (F(1, 50) = 124.08, MSE = 151295, $p < .0001$). For ease of interpretation, example trials and mean RT's corresponding to each condition involved in this interaction are shown in Table 2. A close inspection reveals that the factors involved in the interaction combine to determine whether the order in which the referents appear in the sentence are aligned with the order in which they appear in the graph. Responses were faster for aligned trials than for non-aligned trials. Post hoc Tukey tests revealed that of the 16 comparisons that could be made to test this interpretation of the interaction, 15 were statistically significant ($p < .05$) in the direction predicted.

Error Data. In order that our analysis of error rates might parallel our RTs analysis, we used the same trimming procedure as for the RT data. For each participant we calculated the mean error rates across conditions. A 2x2x2x2 within participants ANOVA was carried out on errors. Mean error rates are shown in Table 3.

Table 2: Sample arrays and RTs from the interaction between Slope, Relational Term and the Match variable in our experiment.

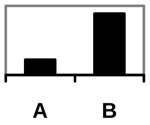
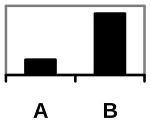
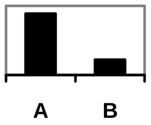
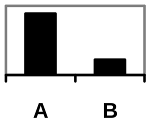
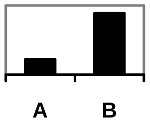
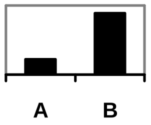
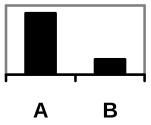
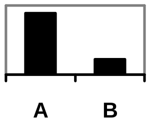
	Ascending		Descending	
	Match	Mismatch	Match	Mismatch
Greater than	 B is greater than A RT = 2313 ms	 A is greater than B RT = 2136 ms	 A is greater than B RT = 1973 ms	 B is greater than A RT = 2387 ms
Less than	 A is less than B RT = 2021 ms	 B is less than A RT = 2429 ms	 B is less than A RT = 2280 ms	 A is less than B RT = 2067 ms

Table 3: Mean percentage errors (standard deviations in italics).

		Ascending		Descending		
Rel-Term	Label Order	Match	Mismatch	Match	Mismatch	
Greater than	Alpha	5.39	3.10	0.49	7.68	4.17
		10.38	*8.65*	*3.50*	*15.71*	
	Non-Alpha	1.47	3.92	1.96	6.21	3.39
		5.94	*9.18*	*6.79*	*14.70*	
		3.43	3.51	1.23	6.95	3.77
Less than	Alpha	5.88	4.58	4.90	3.76	4.78
		12.50	*10.04*	*13.25*	*13.05*	
	Non-Alpha	4.58	3.27	5.39	4.08	4.33
		11.22	*10.42*	*13.52*	*9.63*	
		5.23	3.93	5.15	3.92	4.55

The ANOVA revealed just one significant source of variance - the interaction between Relational term and Match/Mismatch (F (1, 50) = 11.16, MSE = .0079, p < .002). Tests for simple effects revealed that participants made fewer incorrect responses to matching than mismatching trials when the relational term was *greater than* (F(1, 50) = 10.60, MSE = .008, p < .003). However, there was no difference in error rates for matching and mismatching trials when the relational term was *less than* (F(1, 50) = 1.832, MSE = .0089, p > .15). Although we did not predict this interaction *a priori*, differences due to the relational term are predicted by a variety of accounts of how we represent and reason about relationships between objects [for a review see 14].

Given the highly significant three way interaction identified by our analysis of RTs, it is important to note that this interaction does not account for a significant amount of the variance in our error data (F(1, 50) = 2.96, MSE = .0133, p > .05). An examination of Table 3 reveals that the trends present in this interaction do not suggest the existence of a speed-accuracy trade-off in participants’ responses. Although non-significant, the trend is in the opposite direction.

3. Discussion

The first thing that is apparent from our results is that there were significant differences in both the RT and error data which presumably reflect differences in cognitive processing during goal oriented graphical interpretation. We believe that these differences demonstrate that our sentence-graph verification paradigm is an appropriate tool for studying graph comprehension.

The second point to note from our results is that the data indicate that participants do not necessarily represent the relationship shown in a graph using the same relational term as that used in the sentence. This finding is in direct contradiction to the central assumption underlying Clark and Chase’s [11] account of sentence-picture verification. Recall that Clark and Chase argued that there were two separate stages of processing during picture-sentence verification: an initial stage of encoding and a second distinct verification stage where representations of the sentence and graph are compared. Importantly, Clark and Chase argued that the representations of the picture and the sentence should both employ the same relational term to allow direct comparison.

In our experiment, trials where the order of the referents was aligned produced response latencies that differed from latencies for trials where the referents were not aligned. In simple terms, participants read a sentence like *A is greater than B* and then they processed a graphical display in which the referents are either aligned (i.e. A to the left of B on the abscissa), or non-aligned (i.e. in the reverse direction). Aligned trials resulted in shorter response latencies than non-aligned trials. Clearly, if participants made their verification decisions in the manner advocated by Clark and Chase, then they should construct a representation of the graphical display using the same relational term as that used in the sentence. However, if participants did form representations in this way, then we would have expected shorter response latencies when the non-aligned graph matched the linguistic statement (i.e. the magnitude of bar B being less than that of A) than when it did not match (i.e. the magnitude of bar B being greater than that of A). We observed no such difference in response latencies. Consequently, we consider the current data to provide strong evidence against Clark and Chase's suggestion that linguistic and graphical representations are constructed using the same relational term in order to allow a direct comparison.

An alternative explanation that we favour, is that participants employed a strategy whereby they checked the aligned representations for matching relational terms. When the order of the referents was not aligned between descriptions, participants were forced to transform their representation of either the graph or the sentence to allow direct comparison. This transformation could account for the additional cost in processing time for non-aligned trials compared with aligned trials.

Note that Clark and Chase explicitly manipulated the spatial layout of the displays used in their experiments so that participants would first encode the sentence and then verify this description against the picture or vice versa. However, in our study participants were simply told to "verify that the statement is an accurate description of the graphical display". Given this instruction and simultaneous presentation of the descriptions, participants need not necessarily have formed a representation of the sentence prior to verifying it against that of the graph. Although direct comparisons between the current study and the work of Clark and Chase do not require us to be sure that participants in our study encoded the sentence before comprehending the graph, we have recently completed a separate experiment in our laboratory that provides insight into this question. In this experiment participant's eye movements were recorded as they carried out exactly the same sentence-graph verification task [for a similar approach see 15 & 16]. Preliminary analyses of the data from this experiment reveal that on the vast majority of trials participants immediately make a saccade to the sentence in order to read it prior to fixating different portions of the graphical display. We, therefore, assume that in the current RT experiment participants did the same. That is, they constructed a representation of the linguistic statement and subsequently constructed a representation of the graph before making a verification decision.

3. 1 Relevance to Pinker's Account of Graph Comprehension

We will now consider our findings in relation to the account of graph comprehension developed by Pinker [4]. In Pinker's terms our results suggest that the process of interrogating the conceptual message derived from a graph is insensitive to the predicate used to encode the relationship between the referents in that graph. Instead it is driven by the order of the arguments in the proposition.

As mentioned previously, trials where the order of referents in the sentence and the graph were aligned displayed a significant RT advantage over non-aligned trials. This finding is consistent with the claim that the proposition GREATER THAN (A B) in the conceptual message is verified against the proposition LESS THAN (A B) derived from the sentence more quickly than is the proposition GREATER THAN (B A). This effect of alignment further suggests that the graph encoding process is relatively inflexible in that information contained in the sentence does not seem to affect how the graph is initially represented. For example, it appears that a participant who reads the sentence *A is greater than B* and then inspects a non-aligned, descending graph, is unable to prevent themselves constructing a representation in which the relationship is specified GREATER THAN (B A) even though the representation LESS THAN (A B) would be easier to verify. Thus, it appears that certain aspects of the graph encoding process are automatic.

Our results also illustrate that inferential processes are involved during verification. Pinker's use of the term *inference* includes the performance of arithmetic calculations on the quantitative information listed in the conceptual message as well as inferring from the accompanying text what is to be extracted from the graph. The graphical inference for which our experiment provides evidence consists of transformations carried out on representations. As aligned trials have an advantage over non-aligned trials (regardless of relational term, slope or whether the descriptions match), one might assume that the purpose of these transformations is to represent the information from the graph and the sentence so that their referents are in alignment. Once this has been achieved the relationship between the referents specified by each representation may be checked.

Whilst we agree with Pinker that inferential processes are of interest to cognitive psychologists generally, we nevertheless feel that the nature of those inferential processes and the factors which affect them are of considerable interest to any theory of graph comprehension. For example, our finding of an alignment effect is consistent with the claim that participants constructed an analogical representation, such as a mental model [see 17], rather than a propositional description of the graph. They may have then compared this against their representation of the premises (which may also be represented analogically). Although Clark and Chase [11] argued against such an account of their sentence-picture verification paradigm, we have seen that there are good reasons why graph comprehension may differ from picture comprehension. It is conceivable that graphical information may be represented analogically for certain tasks and in certain situations. In the literature on human reasoning [see 14] one tactic used to discriminate between model based [18] and rule based [e.g. 19] theories of deductive inference, has been for each theory to predict what kinds of inferences would be easy under its own representational assumptions. Applying such a tactic to our results, we feel that unless Pinker's propositional account is supplemented with the assumption that people have a preference to build their propositional representation of the graph from left to right, his account cannot easily predict our alignment effect.

3.2 Relation to Other Approaches to Graph Comprehension

Recently, Carpenter and Shah [16] have proposed a model of graph comprehension involving interpretative and integrative processes as well as pattern-recognition. In their model, pattern recognition processes operate on the graph in order to identify its components whilst interpretative processes assign meaning to those component parts (e.g. remembering that an upwardly curving line represents an increasing function). Integrative processes identify referents and associate them with interpreted functions. In a series of experiments where participants' eye movements were recorded whilst they were examining graphs showing complex interactions, Carpenter and Shah claim to have demonstrated that graph comprehension is incremental. That is, people use processes for pattern recognition, interpretation and integration to encode *chunks* of the graph. As the complexity of the graph increases so the cycle of processes is scaled up. Whilst the graphs which Carpenter and Shah investigated were more complex than ours and accordingly were more suitable for their research purposes, their focus is also different from ours. Their interest is in open-ended graph comprehension where we are interested in graph comprehension under constraints. Whilst both methodologies have their strengths and weaknesses, we would argue that graph comprehension is almost always goal oriented.

3.3 Directions for Future Work

Although our work on graphs is in its early stages, possible directions for future work are clear. First, we hope to investigate in more detail the nature of the inferential processes used in graph comprehension. Specifically, we are interested in the transformations performed upon graph representations on non-aligned trials and what those transformations can tell us about graph encoding and comprehension in general. Secondly, we hope to investigate the consequences of inferential processes for graph memory. That is, are trials that require inference prior to verification better remembered than trials that do not? Finally, as we have mentioned above, we have run some experiments where records have been taken of participants' eye movements. These experiments should tell us about the sequence in which processes occur during the extraction of information from graphs. It is hoped that the study of participants' eye movements will shed light on the nature of the inferential processes involved when the referents in the sentence and the graph do not occur in the same order.

4. Conclusions

In this paper we have presented the details of a novel sentence-graph verification technique for the study of how people extract information from graphs. We have argued that this paradigm is analogous to everyday graph comprehension in that it involves goal-oriented rather than open-ended interpretation of graphs. We have described an experiment using this paradigm where it has been demonstrated that the single greatest factor in determining the speed of people's verification response is whether the order of the referents in the graph is the same as the order of the referents in the sentence. We have suggested that this result illuminates some of the micro-processes involved in extracting information from graphs and may be consistent with either an analogical or a propositional explanation of how people represent graphs. Finally, we have outlined some possible directions which future research on this topic might take.

References

1. Zacks, J., Levy, E., Tversky, B. & Schiano, D.: Graphs in Print. In P. Olivier, M. Anderson, B. Meyer (eds.): Diagrammatic Representation and Reasoning. Springer-Verlag, London (In press)
2. Shah, P. & Carpenter, P.A.: Conceptual Limitations in Comprehending Line Graphs. Journal of Experimental Psychology: General (1995) Vol. 124 43-61
3. Lohse, G.L.: The Role of Working Memory on Graphical Information Processing. Behaviour and Information Technology (1997) Vol. 16 297-308
4. Pinker, S.: A Theory of Graph Comprehension. In R. Freedle (ed.): Artificial Intelligence and the Future of Testing. Lawrence Erlbaum Associates Ltd, Hillsdale, NJ (1990) 73-126
5. Kosslyn, S.M.: Understanding Charts and Graphs. Applied Cognitive Psychology (1989) Vol. 3 185-226
6. Gattis, M. & Holyoak, K.J.: Mapping Conceptual to Spatial Relations in Visual Reasoning. Journal of Experimental Psychology: Learning, Memory and Cognition (1996) Vol. 22 231-239
7. Zacks, J. & Tversky, B.: Bars and Lines: A Study of Graphic Communication. Memory and Cognition (1999) Vol. 27 1073-1079
8. Slobin, D.I.: Grammatical Transformations and Sentence Comprehension in Childhood and Adulthood. Journal of Verbal Learning and Verbal Behaviour (1966) Vol. 5 219-227
9. Wason, P.C.: Response to Affirmative and Negative Binary Statements. British Journal of Psychology (1961) Vol. 52, 133-142
10. Greene, J.M.: Syntactic form and Semantic Function. Quarterly Journal of Experimental Psychology (1970) Vol. 22 14-27
11. Clark, H.H & Chase, W.G.: On the Process of Comparing Sentences Against Pictures. Cognitive Psychology (1972) Vol. 3 472-517
12. Gattis, M.: Spatial Metaphor and the Logic of Visual Representation. In AAAI Fall Symposium on Embodied Cognition and Action, Technical Report FS-96-02. AAAI Press, Menlo Park, California (1996) 56-59
13. Lewandowsky, S & Behrens, J.T.: Statistical Graphs and Maps. In F.T. Durso, R.S. Nickerson, R.W. Schvaneveldt, S.T. Dumais, D.S. Lindsay & M.T.H. Chi (eds.): Handbook of Applied Cognition. John Wiley & Sons Ltd, London (1999) 513-549
14. Evans, J.St.B.T., Newstead, S.N. & Byrne, R.M.J.: Human Reasoning: The Psychology of Deduction. Lawrence Erlbaum Associates Ltd, Hove, UK (1993)

15. Loehse, G.L.: Eye Movement-Based Analyses of Graphs and Text: The Next Generation. In Proceedings of the International Conference on Information Systems. Edition 14, ACM (1993)
16. Carpenter, P.A. & Shah, P.: A Model of the Perceptual and Conceptual Processes in Graph Comprehension. Journal of Experimental Psychology: Applied (1998) Vol. 4 75-100
17. Johnson-Laird, P.N.: Mental Models. Cambridge University Press, Cambridge, UK (1983)
18. Johnson-Laird, P.N. & Byrne, R.M.J. Deduction. Lawrence Erlbaum Associates Ltd, Hove, UK (1991)
19. Rips, L.J.: The Psychology of Proof. MIT Press, Cambridge, MA (1994)

Footnotes

1. The order of authors is arbitrary.