Sequential Processing in Comprehension of Hierarchical Graphs

CHRISTOF KÖRNER*

University of Graz, Austria

SUMMARY
Hierarchical graphs represent the relationships between non-numerical entities or concepts (like computer file systems, family trees, etc). Graph nodes represent the concepts and interconnecting lines represent the relationships. We recorded participants’ eye movements while viewing such graphs to test two possible models of graph comprehension. Graph readers had to answer interpretive questions, which required comparisons between two graph nodes. One model postulates a search and a combined search-reasoning stage of graph comprehension (two-stage model), whereas the second model predicts three stages, two stages devoted to the search of the relevant graph nodes and a separate reasoning stage. A detailed analysis of the eye movement data provided clear support for the three-stage model. This is in line with recent studies, which suggest that participants serialize problem solving tasks in order to minimize the overall processing load. Copyright © 2004 John Wiley & Sons, Ltd.

Hierarchical graphs represent the relationships between non-numerical entities or concepts (Butler, 1993). Examples of hierarchical graphs include the structure of computer file systems, tree diagrams, preference information, family trees, and many other sorts of conceptual information.1

Figure 1 shows an example of a hierarchical graph that represents the relationships between a group of people working in a company. The structure of the firm, from the president on down is visualized in the graph.

The usefulness of graphs and diagrams for visualization has often been reported when readers have to learn from texts (e.g. Guri-Rozenblit, 1988; Hegarty & Just, 1993; Mayer & Gallini, 1990; Sweller, Chandler, Tierney, & Cooper, 1990). Educators in mathematics have emphasized how important it is that students learn how to interpret such diagrams as well as produce them from texts (e.g. Barwise & Etchemendy, 1991; Goldin, 1985; Lewis,
Figure 1. A hierarchical graph depicting superior relations between employees of a company. Copyright © 2002 by the American Psychological Association. Adapted with permission.
In addition, Novick and colleagues (Novick, 2001; Novick & Hmelo, 1994; Novick & Hurley, 2001; Novick, Hurley, & Francis, 1999) have emphasized the important role of graphs and diagrams in problem solving and thinking. A considerable amount of research devoted to visualization of conceptual information has been carried out by specialists, such as computer scientists and mathematicians (see visual properties, below). This research activity reflects the valuation of graphical presentation of conceptual information.

A hierarchical graph is accomplished by drawing a labelled node for each concept (employees in Figure 1) and a line connecting two nodes if the pair of represented concepts is in the specified (superior) relation. A line is drawn only between neighbouring nodes, and omitted for all other nodes. The superordinate object is drawn vertically above a subordinate one (directedness). The omission of transitive lines prevents the graph from becoming crowded with connecting lines. The president is, of course, also more senior than the managers’ secretaries, although the lines between the respective nodes have been omitted. The transitive nature of the graph ensures that the reader can infer the ordering of all pairs of concepts. The structure of the relations between the concepts is readily apparent particularly when compared to other possible representations, for example, a list of all the pairs of concepts between which a relation exists. (Note, for example, that the president is in a pairwise relation with all of the other employees.) Not all possible pairs of individuals in the firm are comparable with respect to the hierarchical relation (comparability). For example, the assistant to the president is superior to a secretary, but not to the managers. Such pairs of concepts are called incomparable.

For the proper comprehension of the information contained in hierarchical graphs (or in any pictorial media) a reader needs to establish links or mappings between the represented concepts and their spatial representation in a graph. Gattis (2002a) has identified four kinds of constraints which may guide the mapping of concepts into spatial schemas (see also Gattis, 2002b). Iconicity uses perceptual similarity or resemblance between the represented concepts and their graphic representation to constrain these mappings, associations use similarities between features and properties (without actual resemblance), polarity uses similarities in valence or directionality as constraints, and structural similarity constrains the mappings by exploiting relational structures that are common to the concepts and their graphic representation.

Hierarchical graphs use the upward-downward polarity (directedness) since superordinate concepts are drawn above subordinate ones in the plane. Moreover, they exploit the existing relational structure between the concepts (comparability) and make it visible in the graph. If multiple mappings of concepts into a spatial representation exist they may even conflict, and Körner and Albert (2002a) have investigated the miscomprehension of hierarchical graphs that results when a reader’s knowledge of comparability is replaced by inappropriate, competing knowledge. In the present experiment, however, we imparted the necessary knowledge to participants to ensure comprehension.

Hierarchical graphs are not unique, that is, the same information can be represented by different graphs. Figure 2 shows two different graphs for the same relationships between nine concepts (named a, b, c, etc.). Because there are so many possible ways to draw a graph that represents the same information psychological investigation of what constitutes a comprehensible graph is essential.

Three visual properties for graph presentation are usually discussed in the literature. Planarity refers to whether or not lines in the graph are crossed (Rival, 1993). A hierarchical graph is said to be non-crossed if no line crossings occur; otherwise it is
considered crossed. For example, in Figure 2 the same information is drawn with crossings (right) and without crossings (left). The property of slopes refers to the number of differentially slanted lines in a graph needed to represent the relations between the nodes (e.g. Czyzowicz, Pelc, & Rival, 1990). The graph in Figure 2 (left) has only three slopes, while the graph in Figure 2 (right) includes as many slopes as it includes lines. We call a hierarchical graph upright if it is drawn with as few slopes as possible; otherwise it is called slanted. The property of levels refers to the horizontal adjustment of graph nodes (Pelec & Rival, 1991). For example, in Figure 2 (left) the nodes of neighbouring elements are collected on a horizontal level for each node. In the graph of Figure 2 (right) this is not the case. Therefore, the latter kind of graph is called non-horizontal while the former one is called horizontal. Combining these three dichotomous properties yields eight types of graphs; for example, a type could be crossed, upright, and non-horizontal.

Körner and Albert (2002b) tested the effects of the properties of planarity, slopes and levels on the speed of comprehension in a series of experiments. They presented hierarchical graphs depicting hypothetical preferences between vacation destinations together with an interpretive question (e.g. ‘Is destination a preferred to destination b?’) to their participants. The participants had to verify or deny the question and to press a respective button. An analysis of response latencies showed that responses were faster to non-crossed graphs than to crossed ones. This robust effect of planarity was independent of slopes and levels, which had no effect on the speed of comprehension. In this article we explore at what stage in the course of comprehension planarity exerts its effect; we do so by using eye tracking techniques.

The use of eye tracking is not yet widely spread in graph comprehension research but has already proven useful. Eye tracking data can provide insight into the moment-by-moment processing during graph comprehension by providing an indication of the allocation of attention to different elements of a graph during the course of comprehension or problem solving (see, e.g. Grant & Spivey, 2002). That way, the time-course of comprehension can be segmented and different stages of comprehension can be identified.

For example, Lohse (1997) found individual differences in the time participants spent fixating different areas of displays from graphic decision support software. Similarly, Carpenter and Shah (1998) identified different stages of graph comprehension by analysing the gaze durations on different parts of statistical graphs while participants answered interpretive questions (see also Shah & Carpenter, 1995; Shah, Hegarty, & Mayer, 1999).
What processing might underlie hierarchical graph comprehension? Suppose a participant is asked to verify or deny the question ‘Is object a preferred to object b?’ by means of a hierarchical graph. One subtask of graph comprehension is locating the nodes in the graph that represent the objects mentioned in the question: the target nodes. The other subtask is relating the nodes to each other via the location and connecting lines. One possible sequencing of these processes is that after having located the first target node the participant proceeds by checking to what other nodes it is connected (comparability), and whether those nodes are superordinate or subordinate to the first one (directedness). This formulation would correspond to a two-stage model of graph processing that includes an initial search stage followed by a stage during which search and reasoning processes are carried out simultaneously (e.g. Naglieri, Das, & Jarman, 1990). During such a combined search-reasoning stage the interconnecting lines of the graph are taken into account while search for the second node proceeds. If that were the case a participant would respond to the question more or less immediately after having located the second target node. Also, the second stage of combined search and reasoning would be qualitatively different and take more time than the search alone in the first stage.

Another possibility, however, is a three-stage model that implies two initial search stages and a separated stage during which the graphical reasoning is carried out. More specifically, the first two stages of this model are more or less identical search stages during which the first target node and then the second one is located in the graph. Subsequent to the search stages a reasoning stage occurs during which the relation between the target nodes is checked taking the interconnecting lines into account. For this model of graph comprehension one would not expect an immediate response to the question after location of the second target node; rather, the subsequent reasoning stage must be passed through before a response can be initiated. This should be reflected in a substantial amount of additional eye movements after the second target has been found.

Recent research investigating the relation between visual working memory and eye movements in spatial problem solving suggests that participants use frequent eye movements to serialize a complex task in order to segment it into smaller subtasks; a processing strategy that allows very economical processing (Hayhoe, Bensinger, & Ballard, 1998). Such a view is supported by research that explores linguistically mediated spatial reasoning (Körner & Gilchrist, in press).

In the current experiment we take advantage of the finding that planarity has a substantial effect on the speed of graph comprehension but is independent of slopes and levels, which have no effect (Körner & Albert, 2002b). We directly test the implications of the proposed two and three-stage models. The three-stage model leads to the following clear predictions. First, the time to find (and fixate) the first and second target nodes should be independent of the type of graph. Second, the differences in performance reported by Körner and Albert (2002b) should be a result of an increase in the number of fixations only after the second target is fixated, that is, when information obtained from interconnecting lines is processed. In contrast the two-stage model predicts that the time to fixate the first target node should be independent of the type of graph but the time to fixate the second target node should account for the differences between graph types. In addition following the fixation of the second target node the participant should respond. To test these predictions we carried out an experiment in which we recorded eye movements during a hierarchical graph reasoning task.
EXPERIMENT

Method

Design
The experiment was based on a $2 \times 2 \times 2$ factorial design with repeated measurement on all factors. Visual planarity, visual slopes and visual levels varied on two levels. A graph was either non-crossed or crossed, upright or slanted, horizontal or non-horizontal. We recorded manual responses to the interpretive questions and their latencies. Response latency is defined as the time from graph onset until the participant pressed a button to indicate his/her answer to the question. We also recorded participants’ eye movements during every trial.

Participants
Twelve female students of the University of Graz with an average age of 23.0 years (ranging from 18 to 29 years) participated in the experiment. The participants received ATS 200.– (approximately €14.–) for their participation. They had normal or corrected to normal vision.

Materials
We used eight graphs (the graphs depicted in Figure 2 amongst them), one for each graph type discussed earlier. All graphs represented the same information and differed only with respect to the visual properties of planarity (non-crossed vs. crossed), slopes (upright vs. slanted), and levels (horizontal vs. non-horizontal). The lower case characters a, c, e, o, r, s, u, v, and w were assigned to the nine nodes of each graph.

Four interpretive questions to which participants had to respond ‘yes’ for a correct response (yes questions) and four no questions were constructed. They were all of the form ‘Is a better than b?’ No questions were constructed by asking for incomparable target nodes.

Each graph was paired with each of the eight questions resulting in 64 problems (graph-question combinations). However, for each of the 64 problems a different random assignment of the above characters to the nodes of the graph was used. To increase the power of the data analysis we presented the 64 problems to participants four times resulting in 256 trials.

Apparatus
Two-dimensional eye movements were recorded using an SMI Eye-Link eye tracker (SensoMotoric Instruments GmbH, Germany), which is an infrared video system with a 250 Hz sampling rate and a head movement compensation mechanism. We recorded from both eyes and analysed the data from the eye that produced the best spatial resolution, which was typically less than 0.4°. Displays were presented on a 21” (533.4 mm) monitor with a resolution of $1280 \times 1024$ pixels, while a second PC recorded the eye position data online.

The monitor was adjusted to participants’ eye level at a distance of 60 cm. They were instructed to keep their left hand on the space bar of the keyboard to start a trial. They were further instructed to place the index finger of their right hand on the left-arrow key, which was used for a yes response while placing the middle finger on the adjacent down-arrow key which was used for a no response.
**Task and procedure**

The experiment was divided into two identical sessions with two presentations of the 64 problems per session. At the beginning of a session participants read written instructions that introduced hierarchical graphs as a tool for visualizing preference information. Two example graphs (other than those used in the experiment) were discussed in detail. Participants were told that the graphs illustrated preferences of hypothetical persons for vacation cities and that, to simplify matters, the names of the cities were replaced with single letters. Participants were told that they would have to answer questions concerning preferences visualized in graphs displayed on a computer screen. The concepts of directedness and comparability were explained and illustrated in detail by respective examples. Specifically, we instructed participants to respond ‘no’ to questions involving incomparable nodes. Participants were instructed to respond rapidly and accurately by pressing the appropriate response key (yes or no).

After calibration of the eye tracker participants worked on seven practice trials and two warm-up trials before they entered the main part of the experiment. Before each trial a beep from the computer prompted the participant to start a trial by pressing the space bar with his/her left hand. As soon as the participant had pressed the space bar an interpretive question (e.g. ‘Is a better than c?’) appeared in the centre of the screen for 2500 ms. The screen was cleared and after a delay of 100 ms a fixation cross was presented for 1500 ms at a random position, which was 15° apart from both target nodes. Thereafter, the screen was cleared and the graph was presented for maximally 5000 ms. As soon as the participant had pressed one of the two response keys the screen was cleared and a beep prompting the start of the next trial was played (see Feeney, Hola, Liversedge, Findley, & Metcalfe, 2000, for a similar paradigm). Figure 3 shows the sequence of events of a trial. A small recalibration (drift correction) was performed every five trials. The sequence of 128 problems per session was randomized individually.

**Results**

Data from 3072 trials (12 participants × 256 trials) were obtained. There were 127 (4.1%) incorrect responses, with a range between 1.2% and 15.2% amongst participants. Data from the 2945 correct trials were analysed. The mean manual response time averaged over graphs and questions was 2918 ms (SD = 1334 ms). A $2 \times 2$ factorial ANOVA with repeated measures on the visual properties and the individual mean response times as dependent measure yielded main effects for all the visual properties; however, no interactions were found. The mean response time differences are depicted in Figure 4 (bars). Responses to crossed graphs were substantially slower than to non-crossed graphs (337 ms; $F(1, 11) = 39.83; p < 0.01$). Compared with this effect, the effects of slopes ($F(1, 11) = 6.06; p < 0.05$) and levels ($F(1, 11) = 8.03; p < 0.05$) were comparatively small in size (84 ms and 85 ms, respectively).

For a more detailed analysis based on the eye tracking data we defined a square of the length of 2° visual angle around the letter labelling a node. Only if a fixation was within this area was it counted as a valid fixation of the respective letter and node. In order to test the predictions of the three-stage model we need to partition the sequence of fixations into three phases corresponding to the stages of the model: fixations before and including the first target node fixation; fixations after the first target fixation and including the fixation of the second target; and (if applicable) fixations following the second target node fixation. To this end we first checked whether or not a fixation sequence contained fixations on both
target nodes. This was the case for 2234 of the correct trials (75.9%), which we considered further.

On average, participants made a total of 11.7 ($SD = 5.4$) fixations before they responded. A $2 \times 2 \times 2$ factorial ANOVA with repeated measures on the visual properties and the individual mean number of fixations as dependent measure revealed that participants made 1.62 additional fixations when responding to a crossed graph compared with a non-crossed one, which again resulted in a substantial main effect of planarity ($F(1, 11) = 38.72; p < 0.01$). The differences of 0.30 additional fixations for slanted graphs compared with upright graphs was not significant ($F(1, 11) = 1.81; p > 0.05$), nor was the difference of 0.28 additional fixations for non-horizontal compared with horizontal graphs ($F(1, 11) = 3.24; p > 0.05$). There was a minor planar-by-slope interaction ($F(1, 11) = 5.23; p < 0.05$). Overall, the number of fixations reflected the manual reaction times very well (see Figure 4, line).

Figure 3. Sequence of events in a trial
Segmentation of the fixation sequences showed that the first target node was fixated within 3.65 fixations ($SD = 0.33$), on average; the second one was fixated 3.77 fixations ($SD = 0.60$) later; finally, 4.17 fixations ($SD = 1.62$) occurred after the second target node had been fixated. The substantial amount of fixations of the latter kind (after the second target is found but before a manual response is made) indicates that graph processing continued after both target nodes had been located.

A $2 \times 2 \times 2$ factorial ANOVA with repeated measures on the visual properties and the individual mean number of stage one fixations as dependent measure revealed that the difference of 0.08 fixations for the levels of planarity was not significant; the difference of −0.21 fixations in favour of slanted graphs compared to upright ones was significant, as was the difference 0.20 fixations for levels. There was also a significant three-way interaction (see Table 1).

An ANOVA of the same kind was carried out with the number of fixations for stage two showing effects of the visual property of slopes and a planarity-by-levels interaction. The

Table 1. Analysis of variance of the number of fixations for visual properties and the different stages of the postulated comprehension model

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Stage 1</th>
<th></th>
<th>Stage 2</th>
<th></th>
<th>Stage 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Planarity (P)</td>
<td>1</td>
<td>0.327</td>
<td>2.51</td>
<td>61.37**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slopes (S)</td>
<td>1</td>
<td>6.50*</td>
<td>11.55**</td>
<td>2.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levels (L)</td>
<td>1</td>
<td>5.87*</td>
<td>0.13</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P × S</td>
<td>1</td>
<td>0.033</td>
<td>1.75</td>
<td>3.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P × L</td>
<td>1</td>
<td>4.00</td>
<td>9.33*</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S × L</td>
<td>1</td>
<td>2.87</td>
<td>0.075</td>
<td>7.17*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P × S × L</td>
<td>1</td>
<td>8.60*</td>
<td>0.525</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P × Sub. within group error</td>
<td>11</td>
<td>(0.42)</td>
<td>(0.67)</td>
<td>(0.64)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All reported statistics are within subjects. Values enclosed in parentheses represent mean square errors. Sub. = subjects; *$p < 0.05$; **$p < 0.01$. 

same analysis for stage three revealed that crossed graphs needed 1.28 additional fixations in this comprehension stage compared with non-crossed graphs (main effect planarity). Only a minor slopes-by-levels interaction was found for this stage.

Figure 5 presents the cumulative differences of the number of fixations for the visual properties depending on the different comprehension stages. (Note that the difference cumulated over all stages is identical with the total difference of fixations in Figure 4.)

Three one-way ANOVAs with repeated measures and the cumulative difference of fixation number per stage yielded a moderately significant effect for the visual properties \(F(2, 22) = 5.31; p < 0.05\) for stage one and no effect for stages one and two combined \(F(2, 22) = 1.85; p > 0.05\). As expected, the effect was substantial for stages one, two and three combined \(F(2, 22) = 21.37; p < 0.01\). Planned contrasts showed that the difference of 1.62 fixations for planarity contrasted with the differences of 0.30 fixations for slopes was significant \(F(1, 11) = 29.87; p < 0.01\), as was the contrast between planarity and the difference of 0.28 fixations for levels \(F(1, 11) = 29.03; p < 0.01\).

**DISCUSSION**

The overall response times, as well as the associated number of fixations show that all of the visual properties have an effect on graph comprehension, however to a different extent. Planarity is the dominant visual property; the other properties have a much smaller impact; in fact, no substantial effects for slopes and levels were found with respect to the overall number of fixations (see Figure 4). Moreover, the visual properties are independent of each other.
With respect to the proposed comprehension models, the segmentation of participants’ fixation sequences enabled us to tap into the different phases of graph processing. The first important finding is that the fixation sequence can indeed be partitioned into three major distinct phases using the target node fixations as boundaries. Not only does a third phase exist, it actually involves more fixations than the two previous stages. This is inconsistent with a two-stage model of comprehension that would predict a manual response shortly after fixation of the second target node. It is however, the detailed analysis of the three phases with respect to the visual properties that provides striking evidence for the proposed three-stage model of comprehension. As predicted by the model, the massive impact of planarity on the course of comprehension could be observed only during the last stage; this stage accounts for almost all (1.28) additional fixations that were observed for crossed graphs overall (1.62). Apart from a minor interaction with levels, planarity had no impact whatsoever in the previous stages of comprehension. Rather, we found minor effects of slopes and slopes and levels during those stages. In turn, these visual properties had no effect on stage three of the course of comprehension (apart from a minor interaction).

In sum, these results provide support for the three-stage model of graph comprehension that postulates two search stages followed by a reasoning stage. Therefore, the results suggest that during search for the target nodes in the graph the information that is provided by the interconnecting lines is more or less ignored. At least, it does not affect the search process whether or not those lines are crossed. After all, the presence or absence of crossed lines does not constrain the position of nodes in the graph. However, their location may depend on the horizontal alignment of nodes (levels) and the position to which they are constrained by differently slanted lines (slopes). For that reason, slopes and levels had some minor influence on the search processes but did not affect the following stage.

The effect of planarity suggests that the last stage is a reasoning stage. A possibility to test this assumption more directly is the usage of saccade-contingent display changes. This technique relies on the well-established experimental finding that it is possible to make quite large changes to a display during a saccade which the participant is unaware of (e.g. Inhoff, Starr, Liu, & Wang, 1998). After participants have found the first target node, (i.e. during search for the second target node) a line connected to the first target node can be replaced with a different line in such a way that the hierarchical relationship between the two target nodes is altered. We would expect that the display change does not affect participants’ performance. If the three-stage model is correct, then only location information is acquired during search for the first and second target node while relationship information (coming from interconnecting lines) is obtained in a subsequent (reasoning) stage. A two-stage model would imply that a display change severely disrupts performance on the task. Not only would participants respond slower but they might actually produce a considerable amount of comprehension errors. Experiments that use display changes may also contribute to the idea that reasoning may be grounded in perception (Barsalou, 1999; Goldstone & Basalou, 1998) since the three-stage model implies that there exists a distinct reasoning stage that operates on the results of the previous search stages which are mostly perceptual.

Graph processing in three stages might at first not seem an economical approach to the comprehension task compared with a two-stage process that includes a search and a combined search-reasoning component. Three-stage processing may, however, be computationally less expensive. The visual system can carry out a search for a target object
amongst a set of similar distractors very rapidly and effortlessly (e.g. Duncan & Humphreys, 1989; Treisman, 1988; Wolfe, 1994).

Thus, adding a pure target letter search to the comprehension process may not make a big difference with respect to the overall processing cost. If the location of both target nodes in the graph is known from previous processing a pure reasoning process can focus exclusively on checking the relation between those nodes, and be, in turn less costly than combined search-reasoning. Moreover, prior knowledge of target node locations may constrain the reasoning process to focus on those lines in the graph that are (visually) good candidates of interconnecting lines between the target nodes. In contrast, a combined search-reasoning process would have to be extended to all the nodes that are connected to the only target node processed so far.

The support for three distinct stages of processing, two of which are devoted to search complemented by a final reasoning stage is in line with evidence from other eye tracking studies using spatial reasoning tasks. For example, Hayhoe et al. (1998) observed that specific changes to the spatial display made during saccadic eye movements affected the durations of fixations differently depending on the point in the task when that change occurred. Their result suggests that fixations made in different stages of the ongoing task served different purposes. The reasoning task was broken down into subtasks and the eye movements made were specific to those subtasks. Within this framework each processing stage delivers just the information necessary for the next stage to operate properly on that information, and nothing more. Such an approach can be very economical, yet computationally powerful (Ballard, 1991; Brooks, 1986; Findlay & Gilchrist, 2003). In this sense, the sequential processing approach to graph comprehension may provide the optimal method of information structuring and processing, and the current experiment provides evidence that it is the method adopted by participants.

ACKNOWLEDGEMENTS

Supported by grant 8419 of the Austrian National Bank (OeNB) to the author. I thank Iain D. Gilchrist and the reviewers for valuable comments on earlier versions of the manuscript and Beatrix Ortner for collecting the data.

REFERENCES


