

Building physical and mental models in assembly tasks

Patricia Baggett ^a and Andrzej Ehrenfeucht ^b

^a *School of Education, 610 East University, University of Michigan, Ann Arbor, MI 48109-1259, USA*

^b *Computer Science Department, Campus Box 430, University of Colorado, Boulder, CO 80309-0430, USA*

(Received May 15, 1990; accepted in revised form February 10, 1991)

Abstract

Two videotapes were designed which showed how to assemble an 80-piece object, a lift. The lift had a hierarchical structure (breakdown into subassemblies, subsubassemblies, etc.), which was used as the basis for the units in the videotapes. One tape showed the structure top-down, breadth first, so that actions were not always shown in an executable order. The second combined a top-down approach with a sequential execution of actions. Individuals watched one of the two tapes and then performed the assembly from memory. On a measure of similarity of structure to the correctly built lift, less experienced individuals (females) performed significantly better after viewing the second tape than after viewing the first, while for more experienced subjects (males) there were no differences between the tapes. A principle for instructional designers is suggested: When preparing procedural instructions, combine the task hierarchy and the step-by-step execution as was done here, and less experienced individuals will perform as well as more experienced individuals.

Relevance to industry

The question addressed here is how to design instructions that show and describe a procedure, when the procedure has a hierarchical structure. We show that combining a top-down approach with a sequential execution of actions yields the best performance for inexperienced individuals. The techniques given here should be useful in industry for preparing materials for new or inexperienced workers.

Keywords

Assembly instructions, step-by-step procedures, task hierarchy, procedural structure and actions.

1. Introduction

There is a problem in designing instructions (e.g. a film or videodisc) to show and describe a procedure, when the procedure has a hierarchical structure. The problem is that intuitively one would like to show both the actions of the procedure, in an executable order, and the structure of the task, in as integrated and comprehensible a manner as possible. Then when an individual must perform the procedure from memory, he or she can remember the structure while performing actions, and let the structure guide the performance.

Yet it is not clear whether instructions can be designed that have both these characteristics.

In the study reported here we designed two videotapes showing assembly of an 80-piece object, a lift, shown in figure 1. It is made from the Fischer-Technik assembly kit. When its crank handle is turned, the carrier moves up and down the tower. In each tape, the hierarchical structure (breakdown into subassemblies, subsubassemblies, etc.) was identical. The differences in the two tapes were in what individual shots contained, and in how the shots were sequenced. In the first tape (called non-tree-transformed, or non-TT), the

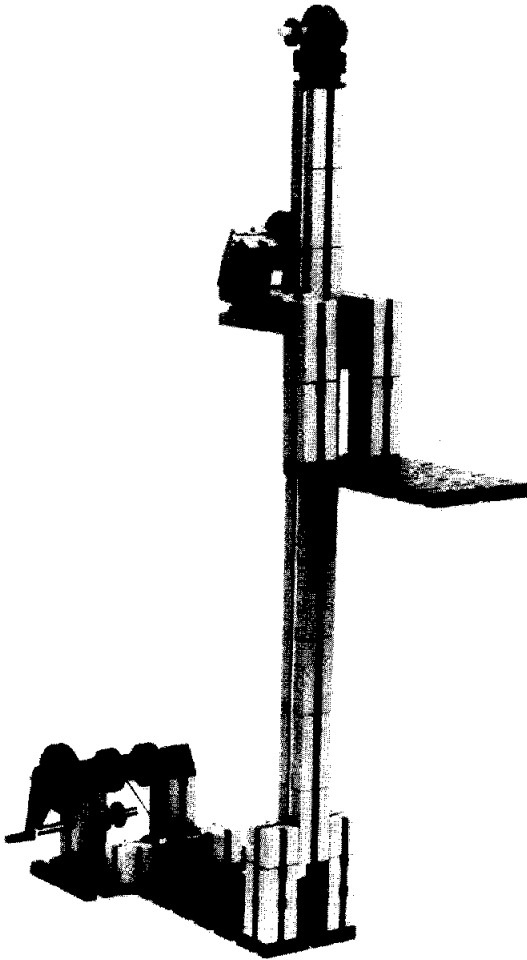


Fig. 1. An 80-piece object, a lift, made from the Fischer-Technik assembly kit.

hierarchy was shown top-down, breadth-first, resulting in a depiction of the top-down structure but a nonsequential execution of actions (to be explained more below). The second (called tree-transformed, or TT) combined a top-down approach with a sequential execution of actions, i.e., it implemented what intuitively seems desirable, as mentioned above. The question we asked was, will these manipulations of sequencing and content lead to performance differences?

Paivio's (1971) dual code theory, hypothesizing that humans store visual and verbal elements separately, seems to suggest that such an ordering should not make a difference. In his theory, the hierarchical structure could be presented top-

down, showing (and describing) actions executed in an order that is impossible for a human (e.g. subassemblies A and B are joined, before A and B have been built). Visual and verbal information showing or describing the structure and the assembly is stored separately, but later mental links are formed to join the visual information into a whole, and the verbal information into a whole, and to cross-index them. Whether the sequential execution information is presented in an order that can actually be executed should not matter.

However, in our framework using a single-code memory with integrated elements from different modalities (described briefly here, but see also Baggett, 1989), we should observe a difference. The group viewing the second videotape should perform the assembly from memory better than the group given the first, and the difference should be greater for individuals less experienced in assembly tasks, because more experienced individuals rely less on instructions. In an earlier study (Baggett, 1984), the importance of temporal synchronization of visual and verbal elements led to a hypothesis of a single-code memory: visual and verbal elements are not stored separately, but together. Concepts consist of multimedia elements. Therefore the question of proper synchronization of elements was a crucial consideration in this study. We hypothesized that simultaneous presentation of hierarchical information and sequentially executable information should lead to an accurate and integrated 'mental model', one that an individual can easily follow in one's head as one performs from memory (see also Gentner and Stevens, 1983; Johnson-Laird, 1983). We will discuss this more below. We first explain the construction of the two videotapes, and then present the results.

2. Representing the structure of an object by a tree

We start with the concept of an ordered, labeled tree. By ordered we mean that the branches coming out from a node are ordered: 1st, 2nd, 3rd, etc. By labeled we mean that each node has a value. As an example, the lift in figure 1 can be represented by the tree in figure 2. There are many possible ways to interpret this tree. For example, A. It is a description of an object. I.e., the lift

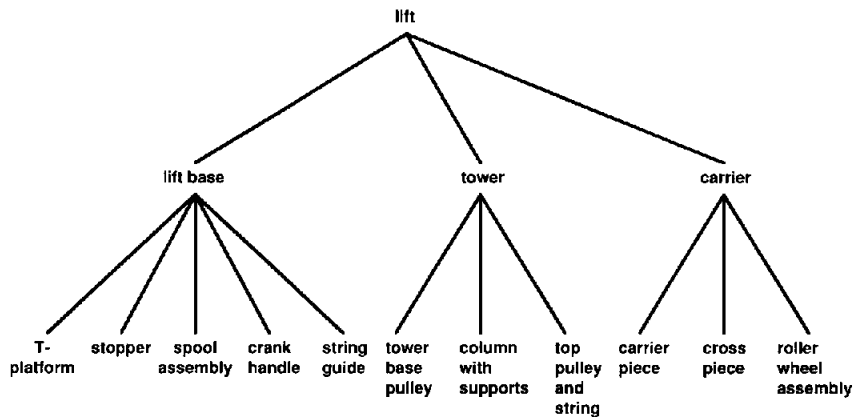


Fig. 2. A hierarchical structure of the lift, showing its breakdown into subassemblies and subsubassemblies, and giving their names.

consists of three subassemblies; the first has five subsubassemblies, etc.

- B. It is a goal structure or mental model. Thus, the goal is to build the lift, and it consists of the following subgoals and subsubgoals...
- C. It is a description of a procedure. Each node is a procedure in which calls are made to nodes directly below, and then assembly actions are performed on the 'returned' subassemblies.

When the tree is viewed as in (C), i.e., a description of a procedure, and we are at the 'lowest' nodes, we get a set of actual pieces, those that make up the subassembly. When we are at a higher node, we construct subassemblies from the pieces, or sub(sub)assemblies below it.

We know that for a given task (or object to be built) we can construct many different trees. In previous work (Baggett and Ehrenfeucht, 1988) we showed that the structure of the tree is important in performance, and we gave the methodology to get the 'right' tree for a given group of people. (We hypothesized (1988) that the tree can be viewed as a person's conceptual structure or mental model of the object: when the person builds the object, he or she mentally breaks it down according to the tree structure and builds it according to the conceptual division. Our data supported this hypothesis.) We also showed (Baggett and Ehrenfeucht, 1990) that the right terminology (names for pieces and subassemblies) is important for helping people match pieces with their names and recall the pieces' names, and we presented a methodology for deriving good terminology. Using those methods, we derived

good structure and terminology to be used in our videotapes for this study. Thus the 'right', or 'typical', tree is shown in figure 2, together with good names for subassemblies. Notice that the lift consists of three major subassemblies, the lift base, tower, and carrier.

The task before us was to construct an instructional videotape for assembling the lift. Obviously a videotape shows information linearly, while the tree we want to display is hierarchical. The question was how to run a video camera over the tree to sequence the shots, and what exactly to show in each shot. (One possibility would be to transform the tree in some way before deciding on the sequence and content of shots.)

We actually constructed two tapes. The first tape began with a fairly extensive demonstration of how the lift functions, pointing out how the string needs to be attached to make the carrier travel along the tower. Then the nodes in figure 2 were shown top-down, breadth first. Even though the structural units were good, the procedure was thus not shown entirely in a step-by-step, executable order. For example, the tape first showed the entire lift, and then its three major subassemblies (lift base, tower, and carrier). It then showed how to put together the three upper-level subassemblies to form the complete lift, before it showed assembly of the lift base or tower or carrier. Next it showed how to assemble individually the lift base, followed by the tower, followed by the carrier. Only one image was shown in each shot, most typically hands assembling a subassembly. At the end, the three were again assembled into the entire

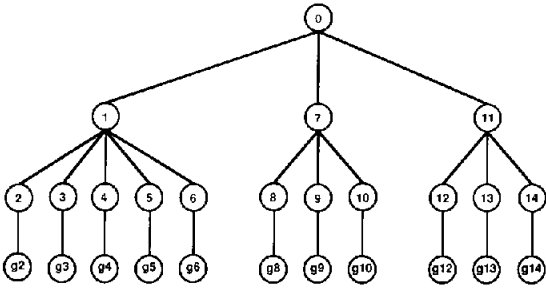


Fig. 3. A hierarchical structure of the lift, with numbered nodes. 'g' means 'get pieces (for a particular subassembly)'. The goal structure is top-down, but actions go the other way.

lift. While the goal structure of the lift was shown, it was not simultaneously integrated with showing actions in an executable order. This content (including the beginning demonstration) can be strongly contrasted to the content in the second tape, as we will describe.

The second tape began with a limited demonstration of how the lift works, and did not point out (during the demonstration) how to attach the string to make the carrier move. Then a top-down, but yet step-by-step procedure for assembly was shown. The following constraints were placed on the order of presentation for the second tape:

- (1) The goal structure (description of the object) needs to be presented top-down: goal, sub-goals, etc.
- (2) The presentation of the actual action (getting pieces, assembling) must be presented in executable order. That is, we do not show how to join two subassemblies before those subassemblies have been built.

In the hierarchy presented in figure 2, the goal structure is top-down, and the actions are the other way (bottom-up); this is more explicitly shown in figure 3. In the figure, the g's indicate groups of pieces that are required for subassemblies. We refer to each g as 'get pieces for the subassembly to be built'. Clearly this goal structure does not allow (1) and (2) to happen simultaneously.

Why do we put the two constraints on the design? They are somewhat arbitrary, but in the case of (1) above, we have indirect evidence that top-down is known to be a good way to design and present algorithms ('top-down development'; see Kruse, 1987; Dijkstra, 1976). And in the case of (2), it seems reasonable to think that if a procedure is not shown in the order in which it is to be performed, individuals, especially those who are less experienced, will get confused about what to do. But the big question is whether the requirements that the instructions be presented top-down and in step-by-step order can be satisfied in the same instructional sequence. To make things more difficult, we add a third constraint: prompting or cueing of the goal. In each node of the tree, the action to be performed is to assemble a subassembly or subassemblies. We will say a node is prompted (or cued) if at least one of the subassemblies is finished (already built) when the node is evoked. (As before, we do not have hard evidence guiding us here to say that this is a good principle, but there are indications that prompting or cueing aids performance. For example, when one is cued with the first part of a passage one is asked to recall, one recalls it better than when not

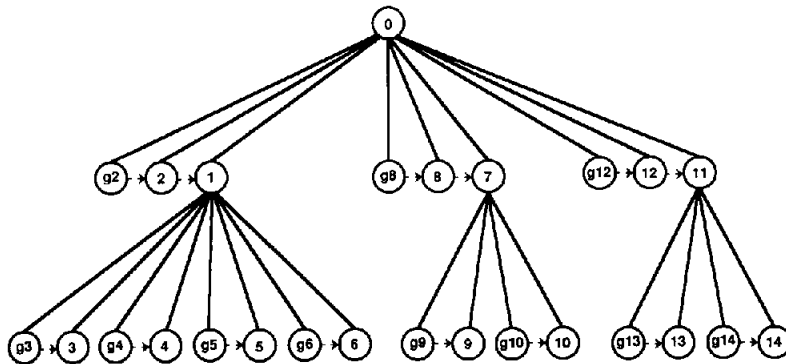


Fig. 4. A tree-transformation of the structure shown in figure 3. Dotted arrows indicate prompting. The subassembly at the head of each arrow is prompted by the item at the foot (left) of the arrow.

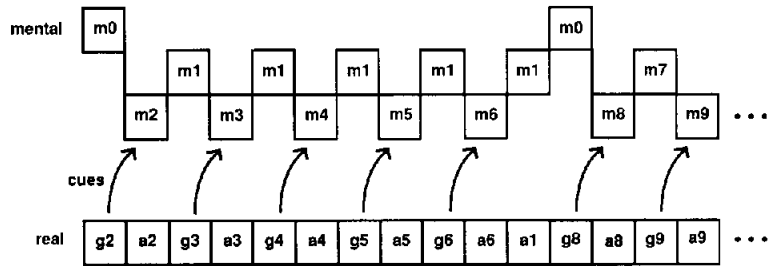


Fig. 5. A representation of the actual shots shown by the two cameras in the tree-transformed tape. Time goes from left to right. The m's stand for mental objects. g means get pieces, and a means assemble subassembly. The numbers correspond to the nodes in figure 4.

given the cue.) The object of prompting in instruction is that the viewer will not be confused about what to do next.

In order to achieve these constraints we did a tree transformation (see figure 4). The transformation is done by removing the leftmost branch from each of the second-level nodes (i.e., removing

nodes 2, 8, and 12), and attaching it directly to the top level node, zero, maintaining the original order. In figure 4 a broken arrow indicates prompting: the (completed) subassembly at the head of each arrow is prompted by the item at the foot (left) of the arrow. The item at the left, the prompt, is typically a 'get pieces' node. In addition, we used

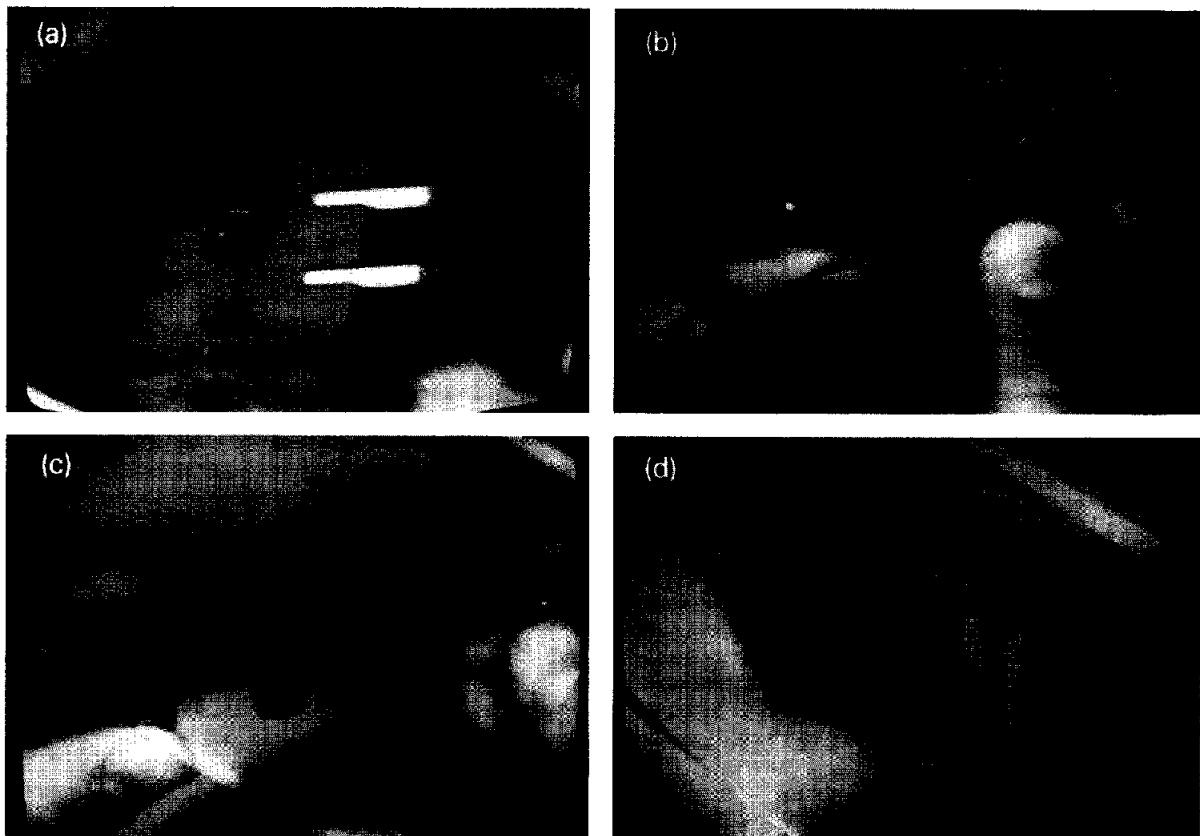


Fig. 6. Actual photos from the non-tree transformed videotape. (a) Assemble T-platform; (b) Assemble stopper; (c) Assemble spool assembly; (d) Assemble lift base.

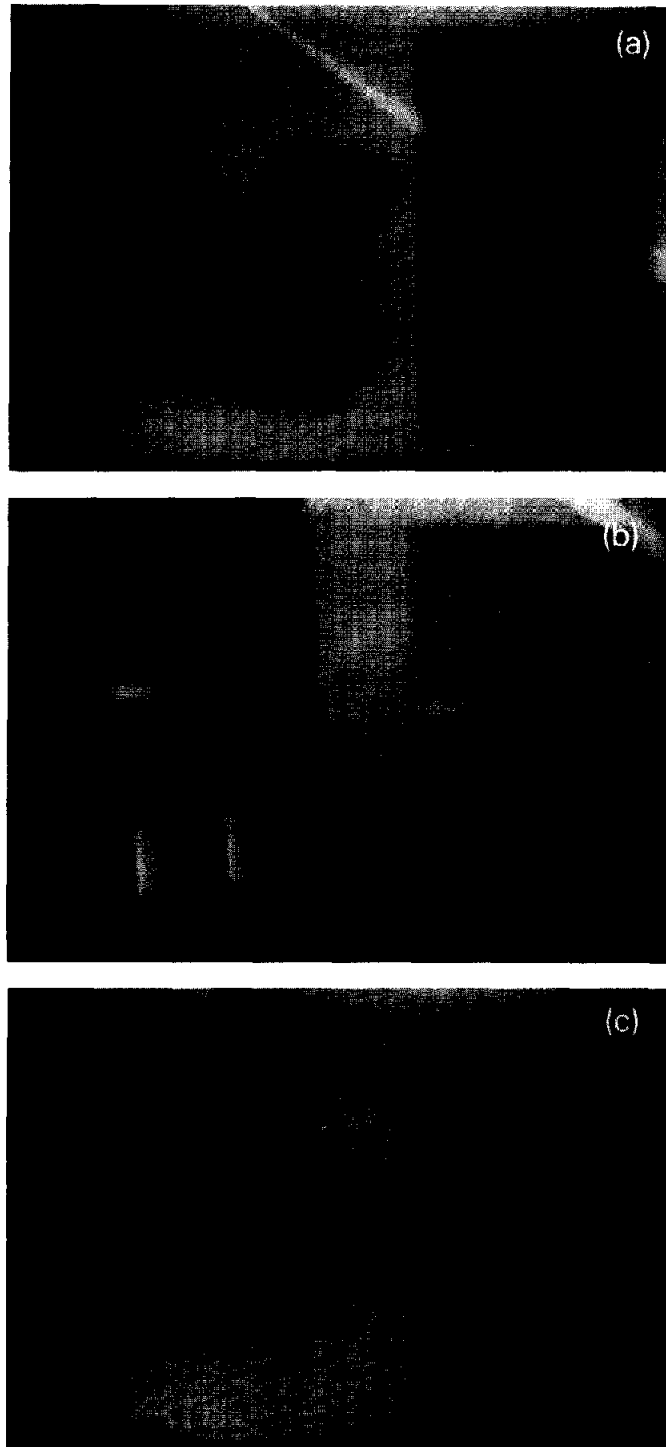


Fig. 7. Actual photos from the tree transformed videotape. Notation for the TT tape corresponds to that used in figures 4 and 5. (a) g2, m0. Get pieces for T-platform. Mental object is lift; (b) a2, m2. Assemble T-platform. Mental object is T-platform; (c) g3, m1. Get pieces for stopper. Mental object is lift base; (d) a3, m3. Assemble stopper. Mental object is stopper; (e) g4, m1. Get pieces for spool assembly. Mental object is lift base; (f) a4, m4. Assemble spool assembly. Mental object is spool assembly.

two cameras, so that there were two images always visible on the videotape: one showed a still shot of the cued, or completed, assembly (which was

meant to be a mental object, as described below), and the other showed hands working toward building that assembly. The external memory aid

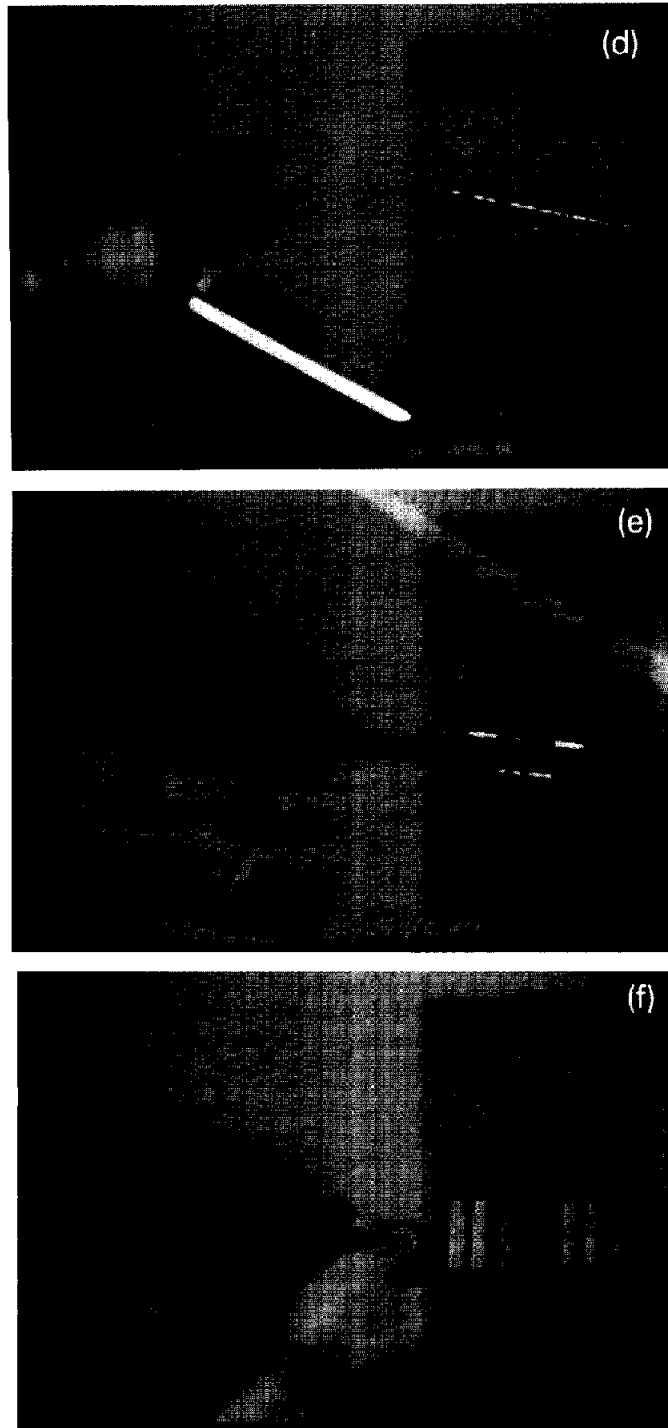


Fig. 7 (continued).

of the image provided by the second camera was meant to keep visible for the viewer the goal (which can be viewed as a mental model of the object).

The idea is to aid a person in building the concept (mental model) of the lift as he or she views the videotape. The person sees first (1) hands laying out a group of pieces, paired with a still image of a 'higher order' goal which includes the subassembly to be made from the pieces. This is followed by (2) hands assembling the pieces shown in (1), paired with the still image of the subassembly formed by just those pieces. The actual schema is shown in figure 5. A 'g' indicates that hands are getting pieces; an 'a' indicates that the pieces are being assembled. An 'm' indicates a mental object, i.e., an object shown in a still image. The sequence of the assembly steps goes from left to right. In the figure, the real action of getting pieces and assembling them occurs on the bottom row. The mental objects are shown in the upper rows. The hypothesis is that when the person sees the mental object, he or she must think, 'I have to build that object'. But it is also explicitly shown on the tape. As the person continues to view the tape, the hope is that after seeing hands getting pieces for a subassembly, and a goal which is a higher-level object, both these elements will go into one (higher-level) subconcept in the viewer's memory. We are attempting, via the sequencing and content given in the shots, to cause the person to build in his or her head the tree which is shown in figure 2, by explicitly providing images to show the structure and to prompt the correct sequence. Then later, when actually building from memory, the hope is that by just seeing his or her own work (after getting pieces and putting them together) the person will be cued or prompted from his or her mental model about what has to be done next. The person can thereby theoretically follow the mental model, concept by concept, that the videotape has carefully laid down.

There is another point to consider as well: the continuity of physical action. Compare figures 3 and 4 (the bases for videotapes one and two respectively), in terms of traversing top-down, with returns. In figure 3 there are long breaks between actions. One must get to the bottom of the tree before one can actually perform an action. (In reality one cannot join two subassemblies until the

two have been built.) In figure 4, every node calls for immediate action. There is no chain of command, and each shot contains at least one executable action. Figure 4 is not as deep.

Still frames from the two videotapes are shown in figures 6 and 7.

3. Method

3.1. Subjects

Sixty-four students from the psychology 100 subject pool at the University of Colorado participated as part of their course requirement, 32 males and 32 females. Most were freshmen and sophomores. Sixteen of each gender were randomly assigned to view each of the two videotapes.

3.2. Experimental design

A 2×2 (videotape (1 = non-TT or 2 = TT) by gender (male or female)) between subjects ANOVA was used. Dependent measures included correctness of assembly from memory, functionality of object assembled, time to work, number of pieces used, and efficiency. These measures will be explained below.

3.3. Procedure

Participants were tested individually. They first filled out a short questionnaire, giving their major, native language, and an estimate of their experience with assembly kits. They were instructed, 'Today you're going to watch a videotape that shows how to build a fairly complicated object from a kit of pieces. After the tape, you'll be asked to build the object from memory, so try to learn as much as you can from the tape'.

Participants then performed a matching task, with the following instructions: 'First, to get you accustomed to the pieces in the assembly kit and their names, please do this matching task. You have before you a collection of one of each of the 48 different pieces in the kit and some sheets of paper containing the 48 names. Spread the sheets out in front of you and put each piece by its correct name. This is not a test. If you have

trouble or want assistance, just ask the experimenter, and she (or he) will help. When you're done, the experimenter will check your matches and correct any you missed'. The matching task took about 5 min.

Participants were positioned before a 15-inch color TV monitor and reminded to learn as much as possible for the memory trial. They placed their chair across from the monitor at any distance they wanted, and the appropriate tape (one or two) was started. Each tape lasted approximately 25 min.

After the tape was completed, they moved to a table to begin the memory trial. They were reminded to build a lift as much like the one they had seen built as possible, and they were told there was no time limit. The experimenter had a kit of pieces, and the subject was required to ask for them one-by-one, either by name or by pointing. The subject had a folder with color photos of the pieces and their names, to aid in getting the pieces. He or she was told that pieces did not have to be used once they were requested; they could be left aside and never used if the subject desired. The experimenter recorded the time the first piece was requested and the time the subject quit, and the order in which the subject requested the pieces.

4. Results and discussion

Performance in building from memory was measured as follows. Abstract graphs of the correctly built lift and of each participant's lift built from memory were drawn. Pieces were represented as nodes and physical connections as links. The number of correct connections in each person's lift (104 were possible) was used as a measure of similarity in structure of the original to a memorial lift. A second performance measure was functionality. It was determined whether or not the lift built from memory contained a handle that, if turned, caused a lifting device to travel up and down a tower. Table 1 presents the results for structure and functionality. A 2×2 ANOVA on the structural measure showed no main effect of gender or of videotape viewed, but a significant interaction ($p < 0.02$). (Figure 8 shows the interaction.) An analysis of simple main effects showed that females performed significantly better on structure ($p < 0.01$) from the tree-transformed

Table 1

Percentage of correct connections (C) and number of functional lifts (F) (16 possible) constructed from memory, as a function of gender of subject and instructional videotape viewed.

Gender		Instructional videotape viewed	
		Non-tree transformed	Tree-transformed
Female	C	34.4	49.4
	F	5	8
Male	C	50.6	47.5
	F	15	9
Average	C	42.5	48.5
	F	20	17

tape, while there were no significant differences in the males' structural performances from the two tapes. For functionality, there was a main effect of gender, with males performing significantly better than females; and an interaction. A simple main effect analysis showed that males performed better on functionality from tape 1 than from tape 2, while females performed equally on both.

Two by two ANOVAs were also performed on time to work during the memory trial and total number of pieces requested. Table 2 gives the means, by gender and by videotape viewed. Neither the main effects nor the interaction were significant for time to work. But for number of pieces requested, females requested significantly fewer than males ($p < 0.02$).

Finally, an efficiency measure was computed as number of correct connections divided by time to perform memory trial. The mean efficiency scores are given in table 3. The 2×2 ANOVA showed no main effects but a significant interaction ($p < 0.05$). However, an analysis of simple main effects failed to reach significance for the difference for females, which was 0.66 for tape 1 and 0.96 for tape 2 (compared to 1.00 and 0.83 for males).

The results show that females equalled males in performance on structure from memory when they first viewed the tree-transformed videotape. When females viewed the non-tree-transformed tape, they performed significantly worse than males. For males, the videotape viewed really didn't matter very much in terms of structural scores; the tree transformed tape did not help (or hurt) males

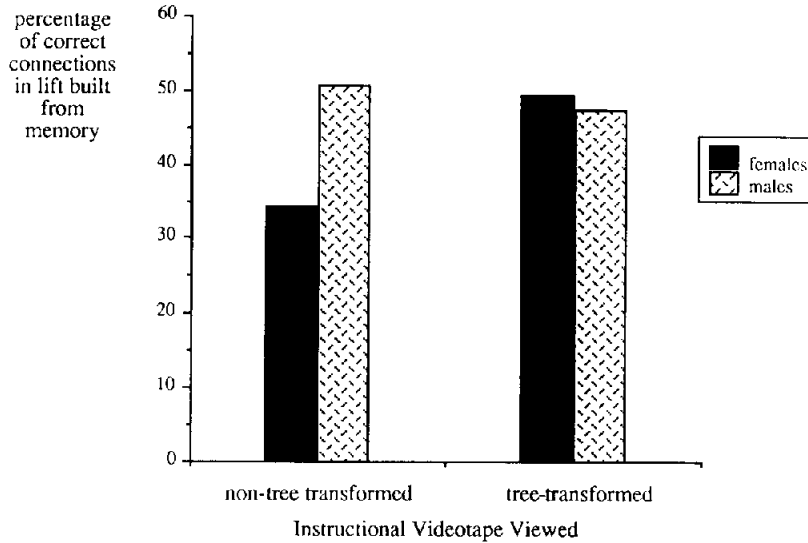


Fig. 8. Percentage of correct connections in lifts built from memory, as a function of gender of subject and instructional videotape viewed. (These data are also presented in table 1.)

Table 2

Mean time (min) to construct lifts from memory (T) and number of pieces requested (P) during the memory trial, as a function of videotape viewed and gender of subject.

Gender		Instructional videotape viewed	
		Non-tree transformed	Tree-transformed
Female	T	60.9	59.1
	P	65	70
Male	T	59.4	70.7
	P	79	77
Average	T	60.15	64.9
	P	72	73.5

Table 3

Mean efficiency (number of correct connections in lift built from memory divided by time to work during memory trial in minutes) as a function of videotape viewed and gender of subject.

Gender	Instructional videotape viewed	
	Non-tree transformed	Tree-transformed
Female	0.66	0.96
Male	1.00	0.83
Average	0.83	0.895

significantly. When males were shown a tape in which a detailed explanation of functionality was shown, they performed better on the functionality measure, whereas females did not.

All participants at the beginning of the session rated their experience with assembly kits, on a scale of 0 (none) to 3 (lots). Mean ratings are given in table 4. The ratios of the male to female ratings for the non-TT group (1.69/1.06 = 1.59) are similar to the ratios of the male to female structural scores (50.6/34.4 = 1.47) for the non-TT tape. So inexperience and the amount of help one derives from the tree-transformed tape are correlated. A possible explanation for why the tree-transformed tape helps inexperienced females perform better is that the organized sequencing and cueing in the video explicitly provides them the

Table 4

Mean ratings of how much subjects had played with assembly kits, on a scale of 0 (none) to 3 (lots).

Gender	Instructional videotape viewed	
	Non-tree transformed	Tree-transformed
Female	1.06	1.31
Male	1.68	1.59
Average	1.37	1.45

opportunity to arrange information as it is presented so that associative links can be formed for good retrieval. No such benefit comes for males, however; their performance from the two tapes is not statistically different.

5. Conclusions

The main considerations in this study were to design instructions for a procedure so that there would be:

- (1) a step-by-step presentation of an executable order;
- (2) simultaneously, a presentation of the hierarchy of goals; and
- (3) visual cueing as a method for creating chains of associations.

The question we asked was, will such a presentation yield better performance from memory than a presentation in which not all three elements are present? Paivio's (1971) stance suggests that individuals are adept at making connections between elements regardless of considerations of juxtaposition. But our stance suggests otherwise. We found that the presentation helps less experienced individuals (females) but plays no role for more experienced individuals. In a previous study (Baggett and Ehrenfeucht, 1988) on assembly, our biggest difference was a gender difference. The finding that a fairly straightforward manipulation of content and sequence can wipe out the difference seems important.

We do not want to overgeneralize the finding, and would like to see it replicated in other tasks. The lift built here consisted of 80 pieces, and one question of interest is whether the result would hold if the task were significantly more or less complex. The manipulation basically allows one to hold less in memory during instruction, and we speculate that this decrease in memory load allows one (especially beginners) to work with and form good chains of associations during learning, regardless of the task's complexity. So we speculate that the result will hold in other situations, al-

though it is possible that the effect may be somewhat dampened and even disappear in highly complex tasks. Of course this is an empirical question which we leave open.

On the basis of the results presented here, we suggest a principle for instructional designers to follow: when preparing procedural instructions, combine the task hierarchy and the step-by-step execution as described above, and novices will perform as well as more experienced individuals.

Acknowledgements

This work was supported by Office of Naval Research contracts N00014-78-C-044, N00014-84-C-0112, and N00014-85-K-0060 in the Cognitive Science Program, Susan Chipman, manager. We thank James Otis, Bud Leonard, and Kate Albers for shooting the video tapes, and many students for helping to test subjects.

References

- Baggett, P., 1984. Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, 76: 408-417.
- Baggett, P., 1989. Understanding visual and verbal messages. In: H. Mandl and J. Levin (Eds.), *Knowledge Acquisition from Text and Pictures*. North Holland, Amsterdam.
- Baggett, P. and Ehrenfeucht, A., 1988. Conceptualizing in assembly tasks. *Human Factors*, 30 (3): 269-284.
- Baggett, P. and Ehrenfeucht, A., 1990. Textual and visual access to a computer by people who know nothing about it. *Proceedings, Sixth International Conference on Systems Documentation, Association of Computing Machinery*; pp. 121-133.
- Dijkstra, E., 1976. *A Discipline of Programming*. Prentice Hall, Englewood Cliffs, NJ.
- Gentner, D. and Stevens, A. (Eds.), 1983. *Mental Models*. Erlbaum, Hillsdale, NJ.
- Johnson-Laird, P.N., 1983. *Mental Models*. Harvard University Press, Cambridge, MA.
- Kruse, R., 1987. *Data Structures and Program Design*. Prentice Hall, Englewood Cliffs, NJ.
- Paivio, A., 1971. *Imagery and Verbal Processes*. Holt, Rinehart, and Winston, NY.