Running head: Procedural learning

Structuring Information Interfaces For Procedural Learning

Jeffrey M. Zacks* and Barbara Tversky†

* Washington University in Saint Louis

† Stanford University

Please address communications to:

Jeffrey M. Zacks

Washington University

Psychology Department

St. Louis, MO 63130-4899

314-935-8454

jzacks@artsci.wustl.edu

Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied, 9*, 88-100.

Abstract

Interface design should be informed by the application of top-down cognitive principles derived from basic theory and research. We applied cognitive design principles from two domains, event cognition and media, to the design of interfaces for teaching procedures. According to theories of event cognition, procedures should be presented hierarchically, organized by objects or large object parts and actions on objects. According to research on effects of media, adding appropriate graphics to text instructions can facilitate learning and memory. These principles were partially supported in two tasks: assembling a musical instrument and building a model. Although both top-down principles were effective in guiding interface design, they were not sufficient. They can be combined with iterative bottom-up methods to produce usable interfaces.

Introduction

Designing Interfaces. There are two ways to design human-computer interfaces: top-down and bottom-up. The top-down method applies general principles derived from cognitive research to a specific interface for a particular task (e.g., Norman, 1988; Shneiderman, 1998). The bottom-up method analyzes the structure of the particular task, varying its' features systematically to determine their optimal values and refining by iterated testing and design (e.g., Egan, Remde, Landauer, Lochbaum, & Gomez, 1995; Nielsen, 1993). The first method has obvious advantages: It takes a small number of general principles derived from basic research on human cognition and applies them to myriad and numerous domains. Thus it both encourages basic research and raises the promise of wide applicability.

Despite these advantages, the top-down method has inevitable limitations and gaps for the design of interfaces in specific cases (Carroll, 1991; Landauer, 1991). General principles are too general to guide the specific design decisions that may ultimately determine the success of an interface. Furthermore, general principles are typically not quantifiable; as such, they do not inform the trade-offs that are an essential part of interface design. These considerations suggest that an ordered hybrid approach is the best: Use the top-down principles to bound the space of possible interfaces and to suggest promising directions; then use the bottom-up approach to refine specific cases.

One common application whose interfaces all too often draw the ire and sighs of users is instructions to assemble an object or operate a device. Such tasks require conveying a continuous sequence of complex actions, and raise questions about the sequence and about how to convey it. Two domains of basic cognitive research may inform interface design for that needy domain: research on event cognition and research on media. Research on event cognition provides information about how people conceive of sequences of actions and research on media provides information about the modalities effective in conveying different kinds of information.

We applied the top-down approach to the design of interfaces for procedural learning. The case we chose to study was object assembly, a task familiar to our population of college students and legendary for inadequate instructions. Designing an effective interface for object assembly has two major components: schematizing the continuous action of assembly and using the media of language and depiction to convey the procedures. The cognitive principles we applied were derived from our own and others' research in two domains: event cognition and effects of various media. We chose two different and representative test applications for these principles, assembly of a musical instrument and assembly of a toy. The interfaces implementing the principles were designed with care, and the detailed evaluation of performance revealed both the benefits and the limitations of the top-down method. The limitations, felicitously, revealed general principles of their own. We measured both immediate performance of the task to be learned and subsequent memory for the steps involved, because although on some occasions it is important simply to be able to complete an assembly task, at other times the task must be learned for later performance.

Event Cognition. Assembly is a common activity in our daily lives, from putting on clothes and fixing a meal to putting together the new equipment in the office or the piece of furniture. Children, too, spend hours assembling things, from nesting cups to Lego palaces. Assembly is a paradigm case of a complex event, an organized sequence of behaviors that has a beginning, middle, and end. Central to events are achievements or accomplishments, such as climbing a mountain or knitting a sweater (see Zacks & Tversky, 2001 for a review). Despite the fact that events are continuous series of actions, people conceive of them as discrete sequences. When asked to segment events into the largest and smallest units that make sense, people do so hierarchically; that is, the large event boundaries coincide more often than chance with the small event boundaries, indicating that people regard the fine units as components of the coarse ones (Newtson, 1973; Zacks, Tversky, & Iyer, 2001). What underlies the boundaries between event segments? People's descriptions of what occurred in each segment provides insight into that. More than 90% of the descriptions were actions on objects, that is, goal-directed behaviors, such as "put on the top sheet" or "rinse the plate" (Zacks et al., 2001). What's more, there were qualitative differences between descriptions of coarse and fine segments. Specifically, large event segments were punctuated by separate objects or large object parts whereas fine event segments were distinguished by different actions performed on the same part or object.

According to the *Principle of Congruence*, external representations such as instructions for assembly should conform to desired internal representations, other things being equal (Tversky, Morrison, & Betrancourt, in press). Applying cognitive event structure to the design of an interface for object assembly insures that the instructions will be compatible with users' mental representations of assembly, enhancing their comprehension. Considerations of cognitive event structure imply that instructions should be segmented and hierarchical and should break the sequence where people do, around objects or large object parts at the coarse level and around articulated actions on objects at the fine level. Hierarchical presentation of instructions has proved to be beneficial in a number of contexts, for example, in comprehending instructions for assembling circuits (Smith & Goodman, 1984), operating machines (Dixon, 1987b), and drawing simple figures (Dixon, 1987a; Dixon, Faries, & Gabrys, 1988).

Media. The media of instructions as well as their structure affect their transparency and have consequent effects on performance. According to the *Principle of Apprehension* (Tversky et al., in press), external representations should be readily perceived and accurately conceived. From general research in cognition, it is known that memory for pictures is superior to verbal presentation of the same material (for a summary, see Paivio, 1986) One account of this widespread phenomenon is that pictorial codes provide more and richer retrieval cues. An additional advantage of pictorial presentation for concrete procedures is that it can directly portray the procedures rather than explaining them in language; it uses depictions of action to convey action. Mapping the spatial to space is compelling and natural in the sense that children and adults in communities all over the world have created such mappings (e.g., Tversky, 2001). In addition, presenting material in two media, pictorial and verbal, is generally superior to presenting material in only a single medium, though of course this depends on the pictorial information being well designed and integrated (Mayer, 2001). In particular, when combining graphics and text it is important that the two be integrated to avoid problems due to splitting attention (Sweller, 1999). Considerable research supports the efficacy of graphic depictions in various applied settings (for reviews, see Levie & Lentz, 1982; Levin, Anglin, & Carney, 1987; Mayer & Gallini, 1990; Winn, 1989). In a recent review, Mayer and Moreno (2002) summarized seven prescriptions for the design of educational interfaces using animation. At the top of their list was the *multimedia principle,* which says that deeper learning results when animation is combined with text.

Animated pictures seem a natural extension of enrichment of stimuli, from words to pictures to moving pictures. Animations have a further possible benefit for assembly in cognitive compatibility: They use change over time to convey change over time. However, animations have disadvantages as well. Animations can be complex, and viewers may not know where to focus attention. This complexity renders them difficult to process and consequently, difficult to remember. Animations are fleeting and cannot be inspected and reinspected as static graphics can, without special interface support and explicit user action. Perhaps for these reasons, and despite their natural correspondence, in practice, animations have not yielded better performance than informationally equivalent static diagrams in a large variety of contexts (see Tversky et al., in press for a review). Nevertheless, because of the natural correspondence of using portrayals of change over time to convey concepts of change over time, hope remains that the proper animation for the proper domain should show benefits.

For conveying events, animated graphics have potential advantages and disadvantages. Although they use change over time to convey change over time, they are continuous, whereas people conceive of events as discrete, so animations do not correspond to the way people conceive of events. On the other hand, animations can convey manner and timing of action, which when complex or subtle, as in knot tying or tennis serves, can be difficult to portray in static graphics.

The top-down approach to designing interfaces for assembly, then, provides two cognitive design principles: First structure the interface to match the structure of the procedure to be taught. This means the interface should explicitly represent the hierarchical structure of the task to be performed. Second, use text and pictures, with a question mark on animated vs. static pictures. We applied these principles to the design of interfaces for two different assembly tasks, assembly of a saxophone and assembly of a toy bug from a construction kit. For the saxophone, we had determined the hierarchical structure from previous research (Zacks et al., 2001), but we double-checked and refined the hierarchy for this experiment. For the toy assembly, we built in the hierarchical structure into instructions. The tasks differ in several ways. Although order of assembly is constrained in both, that is, certain operations must be accomplished before others, there are fewer constraints in the toy assembly.

Experiment 1: Assembling a Saxophone

In the first experiment we taught participants how to put together a tenor saxophone, as if to play it. This activity was chosen because we had normative data regarding how people perceive it and because it is unfamiliar to most of our participants. Although there are minor variations in how one orders the steps involved in this task, we chose one particular ordering as the target for training (see *Table 1*). This small constraint on the naturally occurring activity was intended to capture the common situation in which the affordances of the parts to be assembled permit multiple orders, but only a subset of these are correct. We refer to this as the "Ikea effect," (after the furniture company): the experience of coming to the end of a complex assembly task only to discover that there are still critical parts left over, or that the partially assembled pieces don't fit together because a step was performed too soon. Restricting the order of performance also allowed for precise quantification of how well participants learned order information.

Participants learned from a computer interface that varied on two dimensions. First, we varied the *media* in which the instructions were presented: text, text plus still pictures, or text plus video. Second, we varied the *layout* of the interface, presenting the steps organized hierarchically based on our normative perceptual data or simply as a list. We hypothesized that adding visual depictions to the text descriptions would improve memory for the task instructions, and might thereby improve the quality of their performance of the task. We also hypothesized that structuring the interface in accord with normative conceptual representations would improve participants' ability to learn and remember the temporal order of the task. We predicted that that this would improve their ability to perform the parts of the task in the correct order and later remember that order.

Method

Participants: Thirty-five Stanford University undergraduates participated either as part of an introductory psychology course, or in return for an $8 honorarium. An additional 8 participants were excluded because they indicated they were familiar with woodwind instruments or with the experimental hypotheses (5), or because the equipment failed (3).

Normative analysis of assembling a saxophone. We began with a videotape of a woman assembling a tenor saxophone. This stimulus has been described previously (Zacks et al., 2001) It shows the activity as if viewed from a neutral head-high perspective, about 12 feet from the actor. At the beginning of the movie, the saxophone case is sitting, closed, on a table in the middle of the room. The woman enters, assembles the saxophone, and leaves, at which point the movie ends. The duration is 185 s.

We wanted to construct an interface that presented the activity as perspicuously as possible. To do so, we asked 14 participants drawn from the same population as the main group to watch the movie and divide it into parts. In previous work we have used an on-line measure of event segment perception, in which viewers simply watch a movie and press a button to mark boundaries between meaningful segments (Newtson, 1973; Zacks et al., 2001). For the present purposes it was desirable to have more precisely controlled estimates of segment boundary locations, so we adopted a more intensive procedure, in which observers watched the movie with a controller similar to that on a videotape player, and carefully identified segment boundaries. They then provided verbal descriptions of these boundaries. Each observer divided the activity into seven large parts, and subdivided each of these into three smaller subparts. The numbers seven and three were chosen based median numbers of large and small units produced in the previous online segmentation experiment (Zacks et al., 2001). Two raters (the first author and one other) collated these segment boundaries and descriptions to produce a normative set of large and small units of the activity, then rewrote these descriptions as imperative instructions. The durations and descriptions of each small and large part are given in *Table 1*.

Experimental design and teaching materials. Participants learned to assemble a saxophone from a computer interface. This teaching program was implemented in HyperCard (Apple Computer, Cupertino CA) on a Power Macintosh computer (Apple Computer, Cupertino CA) with two monitors. The monitor to the left had a 21 inch diagonal and resolution of 1280 by 1024 pixels; the monitor to the right had a 17 inch diagonal and resolution of 640 by 480 pixels. The left monitor was used to show the overall instructions, and the right monitor was used to provide demonstrations of the parts of the procedure.

The basic design of the interface is depicted in Figure 1. As noted previously, the interface layout and medium of presentation were varied. In the *unstructured* layout condition, the left screen showed only the small parts arranged in order from 1 to 21, in seven columns of three parts each. In the *structured* condition each large part was identified by listing it separately above its three component small parts. Throughout the experiment, small parts were referred to as “steps,” and large parts as “tasks,” to help make them clearer to the participants. Information was presented in one of three media: *text only*, *text plus still* pictures, or *text plus video*. In the text only condition each part was identified only by providing the text of its description at the appropriate place on the left monitor. This interface offered no interactive features. In the text plus still condition, those text descriptions were accompanied by a small still picture depicting the completion of the part in question. Clicking on a picture caused a detailed, full-screen version of that picture to be displayed on the right monitor. In the text plus video condition the left monitor was identical to the text plus still condition; however, clicking on one of the small pictures caused the movie of the part in question to be played on the right monitor. Clicking on a different picture during the playing of a movie stopped play immediately and started playing of the new movie. Participants could revisit each step as many times as they wished.

Both the layout and the presentation medium were varied between participants. Thus, there were 6 (2 x 3) groups tested. Six participants were tested in each condition except the structured, text condition, in which five were tested.

Procedure. The experiment consisted of five phases. First, participants were exposed to the computer interface. Second, they used the interface to learn how to assemble a saxophone. Third, they assembled the saxophone. Fourth, they were given a surprise memory test for the parts taught by the interface. Finally, they completed a questionnaire that asked about relevant experiences and traits.

After giving informed consent, participants were given a detailed description of the of the procedure. It was explained that they would be learning how to assemble a saxophone, and they would be taught using a computer interface. They were told that after learning the procedure they would be provided with a saxophone and asked to assemble it. First, they were given a brief demonstration of the computer interface, using a set of event parts and movies constructed for a different activity (ironing a shirt) using methods similar to those for the experimental materials.

After this familiarization with the interface, the computer was set up with the saxophone assembly materials and participants were instructed to take as much time as they needed with the interface to learn the task. The experimenter reminded them that they would be asked to carry out the procedure afterwards. When each participant reported that they were ready, the computer was turned off, and a saxophone in a closed case was placed on a table next to the computer. A video camera placed above and behind the table so as to capture all task-related activity was set to record; the presence of the camera was pointed out to the participant. The experimenter sat in the corner of the room and instructed the participant to begin. The experimenter did not intervene in the task unless asked a direct question, in which case he was instructed to give a short, reassuring answer. When the participants reported that they had finished the task, the experimenter turned off the camera.

The experimenter then explained that we would like to test the participant's memory for the information provided in the learning phase. Participants were not informed in advance of the memory test. The left monitor of the computer was set up with an interface that showed an array of rectangles corresponding to the layout the participant had been trained on (structured or unstructured). No pictures were shown, and the right monitor was left blank. It was explained that by clicking in the rectangles and typing. Participants could report their memory for the steps (and tasks, for the structured group) on which they had been trained. Each participant was instructed to recall the text of the instructions as completely and accurately as possible, and was invited to use as much time as was required.

Finally, each participant completed a brief questionnaire, which asked about woodwind instrument use, experience with computer-based learning programs, manual dexterity and mechanical ability, computer literacy, and carefulness in reading instructions. This was given for two reasons: first, to exclude participants who had experience assembling saxophones or related instruments; second, to explore whether individual differences in ability or experience were related to task performance. No relationships were observed between the individual difference measures and the dependent measures, so these will not be discussed further.

Results

We were interested in three aspects of performance. First, how did the experimental manipulations affect participants' performance of the task to be learned? Second, how did the manipulations affect memory for parts of the procedure? Finally, how long did each interface require during training? We conducted analyses focused on each of these questions. For all analyses we adopted a false positive rate of .05.

Performance. Performance of the saxophone assembly task was investigated by having a rater, blind to experimental condition, rate the quality of each participant's performance and the degree to which it was performed in the correct order. Before presenting the details of the rating procedure and the results, let us summarize: There were no statistically significant effects of the experimental manipulations on performance of the saxophone assembly task.

The rater was an undergraduate research assistant who at the time of the experiment was a saxophonist in the Stanford University Band, and who had a number of years experience playing (and, therefore, assembling) a saxophone. The rater viewed each participant's videotaped performance from start to finish, and then reviewed the videotape in detail, providing a structured rating. For each of the 21 small parts, the rater was asked to answer the following questions:

1) Was the part performed? Any attempt to carry out a part was counted.

(Questions 2-5 were conditioned on the answer to question 1 being "yes.")

2) Which part did it follow immediately?
3) How well was it executed. (This was rated on a 1-5 scale, with 1 being poor and 5 being flawless.)
4) Did the participant come back to a part after working for an appreciable amount of time on another part?
5) Did the participant ask the experimenter for help with this part?

After rating all the individual steps, the rater was asked to give an overall evaluation of the participant's performance on a 1-5 scale, with 1 being poor and 5 being flawless. It was stressed that this rating need not simply be an average of the ratings for the 21 steps; rather, the rater was asked to make a holistic evaluation of the performance. In the actual event, the overall evaluations were highly correlated with the mean of part-by-part quality ratings ($r = .87$, $df = 33$, $p < .01$), so we will discuss only the overall ratings. (Results for the part-by-part ratings were equivalent.)

Ratings of performance quality spanned the 1-5 range and were approximately normally distributed, with a mean of 2.83 ($SD = 1.07$, skewness = -.09, kurtosis = -.51). Contrary to hypothesis, there was no statistically significant main effect of layout or media on quality, largest $F(2,29) = 1.75$, *ns*. There was a nonsignificant trend toward an interaction between layout and media, $F(2,29) = 2.92$, *ns*, Cohen's $f = .42$; the pattern of this trend was not clearly interpretable.

To evaluate participants' ability to perform the task in the order specified by the instructions, we calculated how far out of order (if at all) each part was performed. This allowed us to test the hypothesis that the structured layout would facilitate performing the parts of the activity in the correct order. Order errors were calculated as follows: If a part was performed following the correct preceding part (e.g., Part 3 following Part 2) it was given an error score of zero. If a part was performed one step too early or too late (e.g., Part 1 or Part 4 following Part 2) it was given a score of 1, and similarly for larger distances. Based on this method, the mean order error for a participant could range from 0 to 10.52 (assuming performance of all 21 parts). We calculated these mean error scores for each participant and submitted them to an ANOVA. Overall, participants tended to perform the steps in order (94.0% were in the correct location), leading to small scores on the error measure (mean .20 parts, range 0-1.1). There were no significant effects of the experimental manipulations, largest $F = .77$, *ns*, largest Cohen's $f = .23$.

Participants omitted few parts (mean 1.23, range 0-5). Participants rarely restarted parts (mean 1.66, range 0-6). Asking the experimenter for assistance was very rare, occurring on 4 of the 735 parts (21 parts * 35 participants) performed.

In short, none of the measures of task performance showed evidence of being influenced by the experimental manipulations.

Memory. Memory measures, unlike the performance measures, were strongly influenced by both the medium of presentation and the layout of the interface. The memory test was scored by a rater, blind to experimental condition, who was instructed to record which of the 21 small part descriptions were produced by each participant, and in what order. (The large part responses were omitted from scoring to preserve the rater's ignorance of each participant's condition.) The rater was instructed to score each part as correctly reported if one of the participants' responses gave the gist of that part; exact wording was not required. For each of the reported parts, the distance from the correct order of parts was coded exactly as for order of performance (see above). The number of parts reported and mean distance of reported parts were submitted to ANOVAs with layout and media as independent variables.

Participants scored well on the memory test, recalling a mean of 18.5 of 21 parts (range 14-21). As can be seen in Table 2 and Table 3, those trained on the video interface did the best, and those trained on the text interface did the worst. This led to a significant main effect of medium. The video and text conditions differed significantly, $t(21) = 3.3$, $p < .01$, Cohen's $f = .72$. Neither differed significantly from the still picture condition, maximum $t(21) = 1.48$ *ns*. Neither the effect of layout nor the interaction of layout and medium were significantly.

Participants reproduced the order of the small parts with moderate accuracy, leading to a mean order error score on the memory test of .87 small parts (range 0-3.82); that is, parts that were reproduced were on average within one part of their correct location. As can be seen in Table 2 and Table 3, order error was lower on average for the group who studied the structured interface, leading to a main effect of layout which approached but did not reach statistical significance. Neither the main effect of medium nor the layout by medium interaction were statistically significant.

Training time. In any study comparing learning from different media, it is important to examine differences in training time across conditions. Training time may be influenced by how engaging different media are, or how difficult they are to work with. If one condition leads to better memory or performance, but takes longer, it is possible that the improved memory or performance is due to time on task rather than differences in instructional effectiveness.

The time taken to learn the saxophone assembly procedure was automatically recorded by the computer program and submitted to an ANOVA with layout and medium as independent variables. Training times were quite variable, ranging from 96 s to 929 s, with a mean of 372 s (*SD* 213 s). As shown in Table 2 and Table 3, they were longest for participants who studied the video interface and shortest for participants who studied the text interface, leading to a statistically significant main effect of medium. The pairwise difference between the text and video conditions was statistically significant, $t(21) = 3.17$, $p < .01$, Cohen's $f = .69$. The other two pairwise differences were not, largest $t(21) = 1.63$ *ns*. Neither the main effect of layout nor the layout by medium interaction was statistically significant.

Given that medium appeared to have similar effects on study time and on the number of parts recalled, it is reasonable to ask whether the latter was due to the former. We did find study time to be correlated with recall, $r = .35$, $df = 33$, $p < .05$. To further explore this correlation, we calculated a regression model with recall score as the dependent variable that included layout, medium, study time as independent variables, as well as all their interactions. The main effect of medium in this model was significant, $F(2, 23) = 6.0$, $p < .01$. To test the unique contributions of medium and study time, we compared this model to reduced models in which each of the two were removed. This found unique contributions of medium, $F(8, 23) = 1.98$, *ns*, and of study time, $F(6, 23) = 2.48$, *ns* which approached but did not reach statistical significance. Thus, the overall correlation indicates that recall and study time are related, and the regression analysis suggests (weakly) that differences in recall between media may be due to factors above and beyond differences in study time.

Discussion

Participants assembled a saxophone using an interface that was either hierarchically structured according to principles of event cognition or was linearly structured. In addition, the interfaces differed by media: text alone, text plus static pictures, text plus video clips of the assembly steps. Both performance and memory were assessed.

Overall performance, assessed by several measures, was high. Neither manipulating the structure of the interface nor manipulating the medium of the interface produced statistically significant changes in performance. Memory did show effects of interface design. Memory for the components of assembly was better when the presentation was richer, that is, best for text plus video, next for text plus still pictures, and lowest for text alone. This supports the design principle that memory for multimedia is superior to memory for text alone. However, this benefit was accompanied by a cost in training time. The results also gave support for the design principle that the instructional interface should represent the structure of the task to be learned. Recall for order information was improved when the interface organized the instructions in a way that was consistent with the structure of the activity; this difference approached but did not reach statistical significance.

Why did the experimental manipulations influence memory but not online performance? One possibility is that immediate task performance is less sensitive to effects of event structure than memory, because the presence of the object to be assembled and its' particular affordances provide clues to performance that compensate for those absent in the mind. Memory is decontextualized, absent those clues, a more direct measure of the cognitive structures engendered by the interface. For memory, effects were in general correspondence with the cognitive design principles.

The advantage in memory of multimedia presentation over text alone was pronounced, but was accompanied by a substantial increase in study time. Thus, this advantage may partly or wholly due to increased processing time or effort. There was no advantage to animated diagrams over static ones, in corroboration of previous findings.

In sum, the first experiment provided support for effects of both of the top-down design principles on memory for task instructions. Heartened by these results, we turned to a different task: building a model creature using a construction toy. This allowed a conceptual replication in a quite different domain using a different interface design. It also provided a means to relax the rigid scaffolding provided during performance of the saxophone assembly task, where the presence of distinctive objects might provide strong constraints on execution order.

Experiment 2: Building a Bug

In Experiment 2 participants built a model scorpion using a construction toy called Zoob (Primordial, San Francisco CA). Zoob consists of a number of roughly tubular colored plastic pieces that interconnect by ball-and-stick joints and ninety degree snap closures. Zoob has some advantages over the well-known construction toy, Lego (Billund, Denmark). For one thing, it is a new toy, and therefore unfamiliar to our participant population. Next, it has components that vary in shape, size, function, as well as means of connection, rendering it more similar to the objects people typically assemble. We picked assembly of a bug because it has a hierarchical organization of parts: the head and body at the higher level, and their components at the lower level. In contrast to the saxophone, the bug can be assembled in many different orders. However, as in the previous experiment we instructed participants that they should learn to assemble the model in the order depicted in the instructions; again, this allowed us to capture the common situation in which there are task order is important, but is not immediately visible in the partial products of assembly, and allowed for quantification of how well participants learned order information.

Our first aim was to attempt to replicate and extend the effects in Experiment 1 of interface structure on memory for the task, and see whether these could be extended to task performance. As before, we predicted that explicitly representing the structure of the activity in the interface would improve participants' ability to learn and remember the order of the parts of the task, leading to better ability to perform those parts in the correct order and remember their order later. Our second aim was to examine the effect of animation in a context in which manner of execution is especially important. As noted in the Introduction, failures to find an advantage for animation are common and the results of Experiment 1 are thus in good company. In a situation in which the fine motor structure of the activity is especially important, one might expect animation to perform better, because it is able to portray more of that fine structure. Conversely, one might predict the opposite, because the fleeting nature of animation makes it more difficult to study the physical details of a depicted operation, or because animation might encourage mere mimicry rather than real conceptual learning (Palmiter, Elkerton, & Baggett, 1991). We selected Zoob model-building in part because it was identified as especially dependent on information about the fine-grained physical structure of the activity. Because the Zoob pieces fit together in multiple ways (unlike the saxophone parts), manner of assembly is relatively more important than in the previous experiment.

The importance of manner of assembly for this task also required that we eliminate the text-only media condition. Pilot testing indicated that participants simply could not understand brief text descriptions of part assembly with this unfamiliar construction set. (This fact by itself reiterates the value of multimedia for conveying some sorts of descriptions.)

Method

Participants: Thirty-two Washington University undergraduate students (age 18-22, 24 female) participated in return for course credit. An additional 3 participants were replaced due to mechanical problems or experimenter error.

Experimental design and teaching materials. The Zoob construction set used here included six different types of piece, each of which were unique in shape and color. Like other construction toys, Zoob has the advantage of generativity: A very large number of objects can be generated from a small number of pieces. For our purposes, Zoob had the added advantage of being unfamiliar to our participants, having been newly released.

We began with a model of a scorpion described in the graphical instructions accompanying the toy. We slightly altered the target model and constructed a procedure for constructing it, based on those materials. The procedure consisted of 4 large parts, which broke down into 14 small parts (4, 3, 5, and 2 small parts, respectively). We filmed demonstrations of each small and large part. These were filmed from above, showing the actor's arms and the Zoob pieces against a white background. Movies of large parts ranged from 38 to 91 s, and movies of small parts ranged from 9 to 43 s.

We designed a computer interface that represented each part of the activity with a picture showing the result of completing that part (see Figure 2). The interface layout was based on the two-monitor design of Experiment 1, with some modifications. As noted previously, there were only two media used: *text plus still* pictures and *text plus video*. In both conditions, parts were shown as pictures depicting the result of completing that part, on the left monitor. The interface was constructed such that when the participant dragged the computer mouse over a picture, the text of the instruction for that part appeared on the screen, next to the picture. (This mouse-over technique was employed to reduce visual clutter.) In the still pictures condition, clicking on the picture for a part caused a full-screen, more detailed picture to be presented on the right monitor. In the video condition, clicking on the picture for a part caused the video demonstration of that part to play on the right monitor. As in Experiment 1, clicking on another button during a video stopped the current video and started the new one, and each part could be revisited at will.

As before there were two layouts: *structured* and *unstructured*. Because the scorpion model procedure had only 14 small parts rather than 21, we were able to arrange all the small parts on one line at the bottom of the screen. In the structured layout condition, large parts were shown above the small parts as before, with lines on the screen to indicate the grouping of small parts into large parts. For both the structured and the unstructured conditions, a picture of the completed model was shown at the top of the screen. Thus, the structured version showed how the task broke down into 4 large parts, each consisting of 2 to 5 small parts, whereas the unstructured version only showed that the task consisted of 14 small parts. The computer interface was implemented using the same hardware and software as in Experiment 1. Eight participants were run with each of the four combinations of medium (still, video) and layout (structured, unstructured).

Procedure. The procedure followed that of Experiment 1, with one substantive modification. The initial exposure to the computer interface was omitted because pilot testing indicated that the interface for the scorpion task was more intuitive and did not require special training. After giving informed consent, participants were given an overview of the experiment. It was explained that they would be learning how to build a model of a scorpion with a new construction toy, and they would be taught using a computer interface. They were told that after learning the procedure they would be asked to build the model.

Participants were seated at the computer, and the interface was explained. They were instructed to take as much time as needed with the interface to learn the task. The experimenter reminded them that they would be asked to carry out the procedure afterwards.

After each participant completed training, the computer was turned off and the participant was directed to a table with the Zoob pieces in plastic tubs, sorted by type of piece. Each participant was told they should try to build the model exactly as depicted in the computer instructions, performing the parts of the task in the order shown by the instructions. The procedure, including videotaping, was as in Experiment 1.

After completing the model assembly, each participant was given a surprise memory test as in Experiment 1. The interface for the memory test consisted of a set of blank boxes on the left monitor, arranged as the pictures for the learning phase had been. For the unstructured layout condition only boxes for the small parts were included; for the structured condition boxes for small and large parts were included, with a line connecting each small part to its parent large part, as in the learning interface.

Finally, each participant completed a brief questionnaire similar to that used in Experiment 1. This questionnaire asked about familiarity with Zoob and other construction toys, rather than woodwind instrument expertise. None of the participants reported previous experience with Zoob. The questionnaire also asked the same four questions about abilities and experiences relevant to computer-based learning. Again, we observed no relationship between these individual difference measures and the dependent variables, so they will not be discussed further.

Results

As in Experiment 1, we were interested in participants' performance of the model-building task, their memory for the parts of the activity, and the time required to learn the procedure. The analyses took the same form as those for the previous experiment.

Performance. Performance of the model-building task was rated using a procedure similar to that of Experiment 1. Let us provide an overview of the results before describing the details: As in the previous experiment, neither of the manipulations affected performance quality. However, in this experiment there was an effect of layout on performance order: Contrary to hypothesis, participants who studied the hierarchically structured interface were less able to perform the steps in the correct order.

Because the construction toy used here was new, there were no users of it with expertise comparable to that of the saxophonist who rated the saxophone assembly performances in Experiment 1. Therefore, we had two raters code the performances. This allowed us to assess the intersubjective reliability of the coding. Both raters were undergraduate research assistants, and were blind to experimental condition. Each rater scored each of the 14 small parts for whether it was performed, quality of performance (1 to 5), and order of performance, as in Experiment 1, followed by a holistic rating of the overall quality of performance (1 to 5). The two raters nearly always agreed about whether a part was performed (97.8%). They agreed about order of performance for 72.1% of performed parts, and about whether a part was restarted for 75.1% of performed parts. The two raters' judgments of the quality of performance of each step were correlated, $r = .68$, $df = 436$, $p < .01$, and the two were within one point on the scale for 85.8% of parts. These constitute good inter-observer agreement. For all analyses, we adopted conservative criteria for inter-rater agreement. For analyses of how many parts were performed we counted only those parts both raters agreed were performed. For the analyses of order and restarting we counted only those parts where the raters agreed on the order. For part-by-part quality ratings we took the mean of the two raters' judgments.

The two raters' overall evaluations of performance quality were highly correlated, $r = .87$, $df = 30$, $p < .01$. We therefore took as our measure of overall quality the mean of the two raters' scores. As in Experiment 1, overall ratings of performance quality correlated the mean part-by-part quality rating for each participant, $r = .72$, $df = 30$, $p < .01$. Analyses of the overall and part-by-part quality ratings gave equivalent results, so only the former will be discussed, as for Experiment 1. Ratings of overall performance quality ranged from 1 to 4.55 and were distributed throughout that range, with a mean of 2.80 ($SD = 1.11$, skewness = -.15, kurtosis = -1.34). Performance quality did not vary as a function of layout or media; neither main effect nor their interaction was statistically significant in the ANOVA, largest $F(1, 28) = .53$ *ns*.

Performance order error was calculated as in Experiment 1. With 14 small parts, the range of possible mean order error was 0 to 7 (assuming performance of all parts). As we predicted, participants had more difficulty performing the steps of the Zoob task in order than was observed for the saxophone task: 47.6% of parts were out of order, compared to 6.0% for Experiment 1. Mean order errors ranged from zero to 5.0, with a mean of 1.24. As shown in Table 4 and Table 5, order errors were greater for the two structured layout conditions than for the two unstructured conditions. Neither the main effect of medium nor the layout by medium interaction was statistically significant.

Closer examination of the order errors revealed that participants in the structured interface conditions were prone to large displacements of strings of steps. For example, several participants who studied the structured interface began by assembling one of the middle body sections rather than beginning with the exoskeleton as instructed, or reversed the order of assembling the two body sections. We tested this formally in two ways. First, we counted the number of participants in each layout condition who moved at least one step a large distance (greater than three steps). More than half (9 of 16) of those who studied the structured interface did so, whereas only 3 of 16 who studied the unstructured interface did. Second, we counted the number of participants who initiated the model building task with a step other than the first step in the instructions. Of those participants for whom the two raters agreed regarding which step had been performed first, more than half (8 of 15) of those who studied the structured interface began with an incorrect step, compared to only 2 (of 14) for the unstructured conditions. Despite the small sample for count data, both the difference in large displacements and the difference in incorrect starts approached (but did not reach) statistical significance ($\chi^2 = 3.33$, *ns*, and $\chi^2 = 3.33$, *ns*, respectively).

Participants omitted few parts (mean = .69, range 0-5), and rarely restarted parts (mean .53, range 0-6). In this experiment we observed no instances of a participant asking for assistance.

Memory. The memory test was scored using the same method as Experiment 1, but with two raters. Each rater was blind to condition and scored only the small part responses, marking whether each was correctly reproduced, and if so in what order. In this experiment we also asked the raters to note whether a participant referred to a part using incorrect names for any of the components of the model, to obtain a finer-grained measure of verbatim recall. The same two raters that had scored the videotaped performances scored the memory tests. Agreement between the two regarding which parts had been recalled was 93.3%. Of those, the raters agreed about whether the step was performed in the correct order 92.6% of the time, and agreed about whether wrong names were used 87.8% of the time.

As Table 4 and Table 5 show, the number of parts recalled was affected by both the medium and the layout of the interface. In this experiment, those who learned from the still picture interface recalled more parts than those who learned from the video interface. Contrary to our theoretical predictions, the structured interface led to worse recall performance than the unstructured interface. There was no interaction between medium and layout. For those parts were correctly recalled, mean recall order was overall low (mean .63, range 0-2.14). There were no statistically significant effects of the manipulations on recall order, largest $F(1,28) = 2.80$ *ns*.

Training time. The time taken with the interface to learn the procedure was highly variable, mean = 440 s, range 165 s to 1198 s, *SD* 229 s. Somewhat surprisingly, it was not affected by medium, nor by the other manipulations, largest $F(1,28) = 1.04$ *ns*.

Discussion

In this study, participants learned to assembly a toy bug from an interface that was either hierarchically structured or not and that either had text and still pictures or text and video clips of the assembly steps. As before, there were not strong overall effects of interface or medium on performance quality. However, the structured interface yielded *more* order errors in assembly than the unstructured interface. For memory of the component steps without regard for order, the structured interface was again worse than the structured one. Whereas in the previous experiment adding structure to the interface had helped, in this case it hurt. Memory for the instructions was affected by the presentation media as well as interface structure: still pictures were better than video clips. There were no statistically significant effects of either interface structure or interface medium on training time.

Can these findings be reconciled with the cognitive design principles? Let us first consider design principles derived from media. Because the bug task did not permit construction of text-only instructions, all of the interface versions adhered to the multimedia principle of Mayer and Moreno (2002). Although animations have the advantage in this case of providing information about manner of assembly, information that is harder to convey with static graphics, animations were actually inferior to still graphics for memory of the parts of assembly. As noted, there are trade-offs in using animations. The downside of animated graphics is that they are complex and fleeting, and cannot be studied. Properly segmented static graphics can provide adequate information about manner of action as observers can infer the missing information. The present finding, then, is in good company with the many other failures to find benefits of animated over static graphics; it stands out in showing a significant disadvantage to animated graphics, despite the fact they contained more information than the static ones.

The effects of the structure of the interface are subtler to explain. The interface that was structured hierarchically led to more violations of the instructed order and more omission of steps, compared to the interface that whose structure was linear. Recall that assembly of the bug is less constrained than assembly of the saxophone, that several orders are equally plausible. Of those, we selected one. We observed that when constructing the model participants who studied the structured interface were more likely to produce large displacements of a sequence of steps. One possibility is that by explicitly representing the hierarchical displacement of the activity, the structured interface facilitated these wholesale displacements. It may that our hierarchical breakdown of the activity (adapted from that provided by the manufacturer) was infelicitous, and that by calling attention to it the structured interface encouraged participants to deviate from it. At the same time, the hierarchical interface necessarily provides more information, increasing users' cognitive load. This combination of mismatching structure and increased cognitive load may have led to poorer tracking of the steps in the procedure. Given that participants who studied the hierarchical interface were impaired at performing the steps in order, it is unsurprising that we failed to replicate the previous hierarchical advantage for step order in memory. The act of performing steps out of order likely influenced later memory for the instructions.

The build-a-bug task, then, supports the design principle derived from considerations of media, but does not support the design principles derived from event cognition. It does provide a useful boundary condition to applying design principles from event cognition: When the hierarchical structure of a task allows for variation, it may be important to identify an optimal variant in order for structure to be a help rather than a hindrance.

General Discussion

Ideally, design of interfaces should be informed by general principles derived from cognitive psychology. This top-down approach cannot be sufficient for interface design; bottom-up adaptations to particular tasks are also needed. Here, we applied cognitive design principles derived from cognition of events and from effects of media to the design of interfaces for assembly of two objects, a musical instrument and a toy.

Design Principles from Event Cognition. From research on event cognition, it is known that people conceive of events such as assembling an object as discrete rather than continuous, and as hierarchical, organized at the higher level around objects or large parts and at the fine level around actions on the separate objects or object parts. For assembly of the musical instrument, the interface structured accordingly facilitated memory for assembly order. For assembly of the toy, in contrast, the hierarchically structured interface reduced memory. The facilitating effects of the hierarchical interface occurred for the task on which we had independent evidence for event cognition. The interfering effects occurred for the toy, for which the normative event structure came from the manufacturer's instructions, with no guarantee that the hierarchy and order selected matched those of participants. For the toy, order of assembly is less constrained than for the musical instrument, and the intuitions of the manufacturer may not have agreed with those of the users. This underscores the importance of empirically validating such interface choices. The hierarchical interface is more complex than the linear interface, placing more demands on information processing and memory. The added complexity of the hierarchical interface may not have yielded sufficient benefits in comprehension to overcome their increasing demands, especially given that the critical information conveyed seemed to be the manner of attachment rather than the order.

The effects of interface structure were less evident in assembly performance than in memory for the parts and their order. This can be attributed to the nature of the two tasks. Assembly performance is augmented by the actual parts of the object; they themselves have affordances that suggest their manner and order of assembly and they serve as memory cues for assembly. Performance is situated, supported by the presence of the object (Suchman, 1987). Memory, by contrast, is decontextualized, thus more sensitive to accessibility of the information that participants have stored. When order of assembly was critical, as for the musical instrument, hierarchical structure facilitated.

When manner of assembly was critical, as for the toy, hierarchical structure interfered, most likely because of the added cognitive burdens it imposed.

One lesson to be drawn for application of principles of event cognition is that hierarchically structured interfaces are more likely to have effects on memory than on performance. The memory effects suggest that it is likely that a delayed performance task would also show benefits of hierarchical structure . A second lesson is that hierarchical structure is likely to be effective when the primary task of the interface is to convey order of assembly and less likely to be effective when the order is not constrained and the primary task of the interface is to convey manner of assembly.

Design Principles from Media. Considerable research in cognition has shown that presenting two modalities, depictive and descriptive, facilitates memory more than presenting a single modality. The task of assembling the musical instrument found such benefits; the interface presenting pictures and text was superior to that of text alone. This provides support for the dual code and multimedia principles proposed by Paivio (1986) and Mayer and Moreno (2002).

Both tasks compared animated and static graphics. In no cases did we observe statistically significant advantages for animation. This corroborates considerable previous research showing no advantage of animated over informationally equivalent static graphics (Tversky, et al., 2002). Furthermore, for the toy assembly task, animated graphics led to worse memory than static ones. This is probably because of the major disadvantage of animated graphics, that they are complex and fleeting and cannot be reinspected as static graphics can. Notably, the disadvantage of animation occurred in the task where manner of assembly was critical, a case where compatibility considerations would suggest that animations might facilitate. For the saxophone assembly task, animations were studied significantly longer than still pictures, but was not associated with benefits in performance or memory.

The lessons to be drawn for effects of media are clear. First, recall of instructions was improved by adding depictive media to them: In Experiment 1, adding depictions to text instructions improved recall for the text by up to 21%, with a mean improvement of 12% (see Table 2). (However, some of this benefit may have been due to increased study time.) Second, animated depictions had no benefits and in some cases were disadvantageous.

Top-down meets bottom-up. Designing interfaces by first applying general design principles derived from cognitive science has demonstrable advantages. It restricts the realm of possible interfaces, pointing the designer in the right direction by capitalizing on general theories and findings. The top-down approach to interface design proved itself, here, in the common task of assembling objects. General principles from event cognition and from media were helpful in guiding design.

However, the poor showing of the hierarchically structured interface in the Zoob task illustrates the value of the bottom-up approach. The evidence suggested that this interface would have been improved by refining the event decomposition represented by the structured interface. This sort of bottom-up, data-driven design is exactly what was used to arrive at the effective design used in the saxophone interface. This illustrates the general point that the top-down approach is inevitably not sufficient, because it cannot inform the tradeoffs of particular cases, notably, the tradeoff of the value added versus the cognitive costs of extra information. Assessing these in individual cases requires empirical study and iterative design methods (Landauer, 1995; Nielsen, 1993).

Data-driven design has limitations as well. In the present context, evaluation of performance quality and memory accuracy require laborious coding of video and recall data, placing severe restrictions on statistical power. The practical implication of this is that methods for rapid prototyping and evaluation of necessary to solve real design problems. These limitations also raise an important caveat We would discourage readers from drawing conclusions from the failure to detect any particular effect in these studies, given the relatively small samples.

The essential finding is that interface design can benefit from the application of cognitive design principles, even when the resulting interfaces are more complex, placing greater cognitive demands. More generally, basic research on event cognition and on media provide effective guidelines for design of interfaces. Principles derived from research on event cognition and on media yielded interfaces for assembly that were, on the whole, successful. Where they were not, the outcomes yielded insights that also inform interface design. Specifically, the interface designed discretely and hierarchically in correspondence with research on event cognition facilitated memory for assembly of the musical instrument but interfered with memory for assembly of the toy. The explanation lies in the challenges of each of the assembly tasks; for the task where hierarchical structure facilitated, order of assembly was obligatory whereas the task where hierarchical structure interfered afforded several different assembly orders. Design principles are effective in guiding interface construction, but can be much more powerful when combined with sound empirical iterative design.

References

Carroll, J. M. (1991). Introduction: the kittle house manifesto. In J. M. Carroll (Ed.), *Designing interaction: psychology at the human-computer interface* (pp. 1-16). Cambridge: Cambridge University Press.

Dixon, P. (1987a). The processing of organizational and component step information in written directions. *Journal of Memory & Language, 26*(1), 24-35.

Dixon, P. (1987b). The structure of mental plans for following directions. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13*(1), 18-26.

Dixon, P., Faries, J., & Gabrys, G. (1988). The role of explicit action statements in understanding and using written directions. *Journal of Memory & Language, 27*(6), 649-667.

Egan, D. E., Remde, J. R., Landauer, T. K., Lochbaum, C. C., & Gomez, L. M. (1995). Behavioral evaluation and analysis of a hypertext browser. In R. M. Baecker, J. Grudin, W. A. S. Buxton & S. Greenberg (Eds.), *Readings in human-computer interaction toward the year 2000* (2nd ed., pp. 843-848). San Francisco: Morgan Kaufmann Publishers.

Landauer, T. K. (1991). Let's get real: a position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In J. M. Carroll (Ed.), *Designing interaction: psychology at the human-computer interface* (pp. 60-73). Cambridge: Cambridge University Press.

Landauer, T. K. (1995). The trouble with computers: Usefulness, usability, and productivity (pp. xiii, 425). Cambridge, Mass.: MIT Press.

Levie, W. H., & Lentz, R. (1982). Effects of text illustration: a review of research. *Educational Communication and Technology Review, 30*(4), 195-232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration I: Basic research* (pp. 116-135). New York: Springer.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge ; New York: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715-726.

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87-99.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28-38.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Paivio, A. (1986). *Mental representations: a dual coding approach*. New York: Oxford University Press.

Palmiter, S. L., Elkerton, J., & Baggett, P. (1991). Animated demonstrations vs written instructions for learning procedural tasks: A preliminary investigation. *International Journal of Man-Machine Studies, 34*(5), 687-701.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer interaction* (3rd ed.). Reading, Mass: Addison-Wesley.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition & Instruction, 1*(4), 359-396.

Suchman, L. A. (1987). *Plans and situated actions: the pattern of human-machine communication*. Cambridge: Cambridge University Press.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Victoria, Australia: Acer Press.

Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79-112). Cambridge, MA: The MIT Press.

Tversky, B., Morrison, J. B., & Betrancourt, M. (in press). Animation: Can it facilitate? *International Journal of Human Computer Studies.*

Winn, W. (1989). The role of graphics in training documents: toward an explanatory theory of how they communicate. *IEEE Transactions on Professional Communication, 32*(4), 300-309.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29-58.

Author Note

We would like to thank Brady Beaubien, Jonathan Bobb, Mary Dowd, Andrea Gallagher, Crosby Grant, Frederica Mayer, Matt Sorem, and Margaret Sheridan for assistance in data collection. Thanks to The Starving Musician (Mountain View, CA) for providing saxophones, and Primordial, LLC (San Francisco, CA) for providing Zoob, for use in this research. This research was supported by a grant from the Phi Beta Kappa Society, Northern California chapter, NIH grant # 1RO3MH62318-01 to Washington University and Office of Naval Research, Grants Number NOOO14-PP-1-O649, N000140110717, and N000140210534 to Stanford University. A portion of these data were presented at the 1999 meeting of the Psychonomic Society.

Table 1

Description of the parts of assembling a saxophone, derived from 14 undergraduate observers. (The duration in the stimulus movie of each step, in seconds, is given in parentheses.)

Large part	Small part
Take out the saxophone. (32.5)	□
	Open the saxophone case. (18.8)
	Take out the saxophone body. (7.4)
	Take out the swab. (6.3)
Clean the saxophone. (40.2)	□
	Pick up the cleaning cloth. (15.7)
	Wipe the saxophone body with the cleaning cloth. (23.5)
	Put the cleaning cloth in the case. (1.0)
Attach the neck. (14.7)	□
	Pick up the neck. (3.0)
	Insert the neck into the saxophone body. (7.2)
	Tighten the neck screw. (4.6)
Attach the neckstrap. (21.3)	□
	Put on the saxophone neckstrap. (97.3)
	Adjust the fit of the neckstrap. (4.7)
	Attach the neckstrap to the saxophone. (6.7)
Attach the mouthpiece. (15.0)	□
	Pick up the saxophone mouthpiece (4.4)

	Attach the mouthpiece to the neck. (7.9)
	Put the mouthpiece cover in the case. (2.7)

(Table 1, continued.)

Attach the reed. (38.8)	
	Wet the reed in your mouth. (14.7)
	Insert the reed between the mouthpiece and the ligature. (17.1)
	Tighten the ligature screws. (7.0)
Put down the saxophone. (13.7)	
	Close the saxophone case. (5.1)
	Put the saxophone down on the case. (6.4)
	Leave the room. (2.2)

Table 2

In Experiment 1, amount recalled and training time were affected by medium, and recall order was improved by adding structure to the interface.

		Text	Still pictures	Video
Number of parts recalled				
	Structured	16.2 (1.24)	19.33 (0.99)	19.67 (0.62)
	Unstructured	18 (0.58)	17.83 (0.87)	19.67 (0.49)
Recall order error				
	Structured	0.29 (0.06)	0.71 (0.42)	0.57 (0.31)
	Unstructured	1.40 (0.50)	1.66 (0.59)	0.50 (0.20)
Training time (s)				
	Structured	253 (99)	335 (49)	516 (84)
	Unstructured	236 (63)	422 (118)	452 (71)

Note: Values are means, with standard deviations in parentheses.

Table 3

ANOVAs for amount recalled, recall order, and training time in Experiment 1.

	Source	*df*	*MSE*	*F*	*Cohen's f*
Number of parts recalled					
	Layout (L)	1	0.01	0.00	0.01
	Media (M)	2	17.77	4.62*	0.53
	L x M	2	7.78	2.02	0.33
	within-group error	29	3.85		
Recall order error					
	Layout (L)	1	3.73	3.97	0.34
	Media (M)	2	1.28	1.36	0.28
	L x M	2	1.19	1.26	0.27
	within-group error	29	0.94		
Training time					
	Layout (L)	1	215.01	0.01	0.01
	Media (M)	2	165784.00	4.11*	0.52
	L x M	2	18051.40	0.45	.016
	within-group error	29	40317.44		

* $p < .01$

Table 4

In Experiment 2, performance and memory were affected by memory, and memory was effected by media.

		Still pictures	Video
Performance order error			
	Structured	2.10 (0.53)	1.43 (0.42)
	Unstructured	0.51 (0.20)	0.92 (0.24)
Number of parts recalled			
	Structured	12.95 (0.51)	10.55 (1.33)
	Unstructured	13.75 (0.16)	12.86 (0.43)

Note: Values are means, with standard deviations in parentheses.

Table 5

ANOVAs for performance order and amount recalled in Experiment 2.

	Source	*df*	*MSE*	*F*	*Cohen's f*
Performance order error					
	Layout (L)	1	8.79	7.91**	0.51
	Media (M)	1	0.13	0.12	0.06
	L x M	1	2.35	2.13	0.24
	within-group error	28	1.11		
Number of parts recalled					
	Layout (L)	1	0.10	4.32*	0.36
	Media (M)	1	0.11	4.80*	0.38
	L x M	1	0.02	1.02	0.17
	within-group error	28	0.02		

* $p < .05$

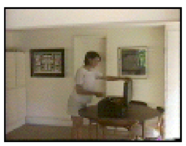
** $p < .01$

Figure Captions

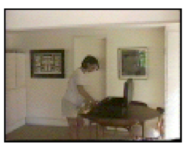
Figure 1. A computer interface for teaching how to assemble a saxophone (Experiment 1). Shown is the layout of the left of two monitors; the right monitor was used to present still pictures or video demonstrations. The 21 small parts of the procedure were arranged across the screen, top to bottom and left to right in rows of 3. In the structured version of the interface (shown here) each of the 7 large parts was depicted above the 3 small parts that made it up, with a larger picture and font; in the unstructured version these were omitted. Pictures were present in the still picture and video conditions; in the text only condition these were omitted. In the still picture condition, clicking on one of these pictures caused a larger, more detailed picture to be shown on the right monitor. In the video condition, clicking on a picture caused a video demonstration of that part to be shown on the right monitor.

Figure 2. A computer interface for teaching how to build a model scorpion (Experiment 2). Shown is the layout of the left of two monitors; the right monitor was used to present still pictures or video demonstrations. The 14 small parts of the procedure were arranged across the bottom of the screen in a single row. In the structured version of the interface (shown here) the small parts were grouped into 4 large parts, shown above and connected by lines. In the unstructured version the large parts and connected lines were omitted. In both the structured and unstructured conditions the picture of the completed model, representing the task as a whole, was shown at the top. In the still picture condition, clicking on one of the pictures (except the top picture) caused a larger, more detailed picture to be shown on the right monitor. In the video condition, clicking on a picture caused a video demonstration of that part to be shown on the right monitor. As can be seen for the first large part, dragging the mouse (hand cursor) over the picture for a part caused the text description of that part to be shown.

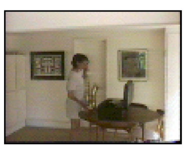
Take out the saxophone.



Open the saxophone case.

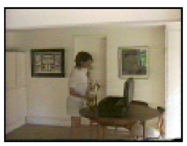


Take out the saxophone body.

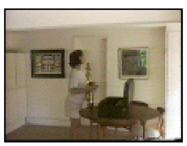


Take out the swab.

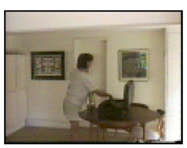
Clean the saxophone.



Pick up the cleaning cloth.

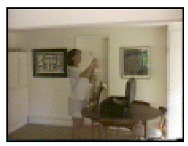


Wipe the saxophone body with the cleaning cloth.

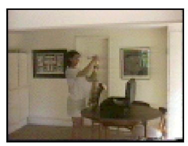


Put the cleaning cloth in the case.

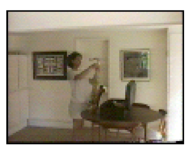
Attach the saxophone neck.



Pick up the neck.

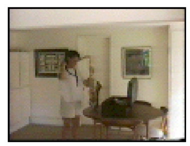


Insert the neck into the saxophone body.

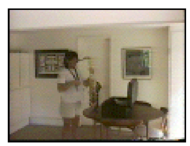


Tighten the neck screw.

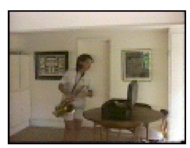
Attach the saxophone neckstrap.



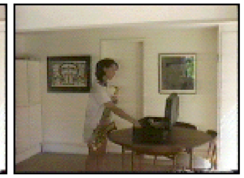
Put on the saxophone neckstrap.



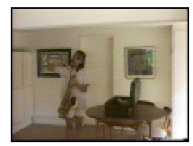
Adjust the fit of the neckstrap.



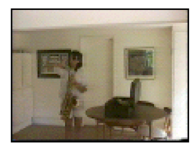
Attach the neckstrap to the saxophone.



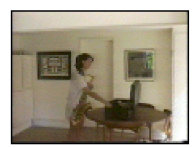
Attach the saxophone mouthpiece.



Pick up the saxophone mouthpiece.

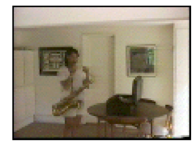


Attach the mouthpiece to the neck.

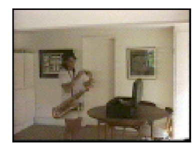


Put the mouthpiece cover in the case.

Attach the saxophone reed.

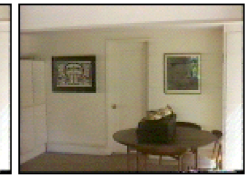


Wet the reed in your mouth.

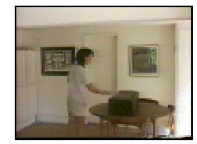


Place the reed on the flat part of the mouthpiece and slide the ligature over it.

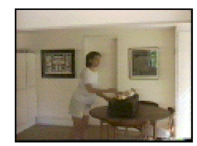
Tighten the ligature screws.



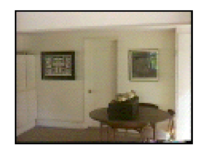
Put down the saxophone.



Close the saxophone case.



Put the saxophone down on the case.



Leave the room.

Build the exoskeleton.