Identification and Validation of Cognitive Design Principles for Automated Generation of Assembly Instructions

Julie Heiser[1], Doantam Phan[2], Maneesh Agrawala[3], Barbara Tversky[1], and Pat Hanrahan[2]

[1]Stanford University
Dept. of Psychology,
Stanford, CA 94305

[2]Stanford University
Dept. of Computer Science,
Stanford, CA 94305

[3]Microsoft Research
One Microsoft Way,
Redmond, WA 98052

{jheiser, bt} @ psych.stanford.edu, {dphan, maneesh, hanrahan} @ graphics.stanford.edu

ABSTRACT

Designing effective instructions for everyday products is challenging. One reason is that designers lack a set of design principles for producing visually comprehensible and accessible instructions. We describe an approach for identifying such design principles through experiments investigating the production, preference, and comprehension of assembly instructions for furniture. We instantiate these principles into an algorithm that automatically generates assembly instructions. Finally, we perform a user study comparing our computer-generated instructions to factory-provided and highly rated hand-designed instructions. Our results indicate that the computer-generated instructions informed by our cognitive design principles significantly reduce assembly time an average of 35% and error by 50%. Details of the experimental methodology and the implementation of the automated system are described.

Categories and Subject Descriptors

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. Evaluation/methodology.

General Terms

Design, Experimentation, Human Factors, Algorithms

Keywords

Diagrams, design principles, visual instructions, assembly instructions, spatial ability.

INTRODUCTION

Making everyday products, such as furniture, appliances, and toys user-friendly is a challenge, especially as the complexity of the products increases [11, 14]. A critical element of usability is the assembly instructions provided by the manufacturer. These instructions commonly use diagrams to convey structural or functional relations between the components of the product. Well-designed instructions emphasize the important aspects of the instructions while de-emphasizing or removing unnecessary details. Such instructions are sensitive to the constraints of human information processing systems, including perception and visual reasoning. Unfortunately, principles for designing effective visual instructions that are informed by human cognition are lacking. As a result, instructions are often convoluted and do not convey information users need. Empirically analyzing the production, comprehension and use of visual instructions can provide not only an understanding of how we interact with such information, but also guidelines for the creation and automation of visual instructions.

In this paper, we describe an approach to identify and validate cognitive design principles for creating effective assembly instructions. We integrate empirical research on human cognition and perception of assembly instructions with the development of algorithms in an automated instruction design system. We focus on the domain of furniture. Like appliances, toys, and other consumer products, furniture has parts that must be configured appropriately in order to function properly. We believe that the cognitive design principles identified for furniture assembly instructions can be generalized to these other products, and that our approach to automatically generating effective visualizations can be generalized to other domains.

We identify and validate cognitive design principles for furniture assembly in five stages.

1. **Production:** Users assemble a TV stand using only a photograph of the assembled product as a guide. These users then produce a set of instructions.
2. **Preference:** A second group of users rate the instructions produced by users in Stage 1. We analyze the high-rated instructions for common characteristics. Design principles are formulated which emphasize these characteristics.
3. **Comprehension:** We test the design principles for efficacy by having a third group of users assemble the TV stand with instructions that have characteristics given high ratings by users in Stage 2. We collect feedback from the users on the elements of the instructions they found to be helpful or confusing, and revise the design principles.
4. **Instantiation:** We instantiate the most effective design principles algorithmically within an automated assembly instruction design system.
5. **Usability:** We conduct a user study to validate the cognitive design principles and the instructions generated by the computer algorithm in Stage 4 by having users assemble the TV Stand with our computer-generated instructions.

The fives stages follow a top-down approach to interface design in which the experimental work guides the design of the automated system. We first present a series of human-subject experiments we conducted in Stages 1, 2 and 3 to identify a set of cognitive design principles for creating effective assembly instructions. Although we have already presented the details of

AVI '04, May 25-28, 2004, Gallipoli (LE), Italy

Stage 4, the algorithmic instantiation, in Agrawala et al. 2003 [2], for completeness we briefly summarize the approach here[1]. Finally, in Stage 5, we test the usability of the instructions generated by our system and we find that the computer-generated instructions significantly improve performance time by an average of 35% and reduce error rates by an average of 50% on a furniture assembly task.

RELATED WORK

The process of using empirical research to derive principles for the design of visualizations is an approach taken by researchers in cognitive science and human-computer interaction. The five-stage approach described here has been applied to design problems in other domains besides assembly instructions. Agrawala and Stolte employed a similar approach for automating the design of route maps [1, 8, 19, 20]. Mackinlay encoded graphic design criteria in a system to enable the automatic presentation of two-dimensional relational information [10]. Seligmann and Feiner developed an intent-based illustration system (IBIS) to generate diagrams that communicate the location and other properties of objects in an equipment maintenance and repair system [16].

Research in cognitive psychology has shown that well-designed instructions can effectively convey the structure of an object and the spatial relations among its components. Effective instructions use a variety of diagrammatic techniques to emphasize important aspects of the instructions and remove unnecessary details. For example, extra-pictorial devices, such as arrows and guidelines [21], convey actions and sequences. Graphic techniques for facilitating visual search, such as the use of color, can improve the operation of a system by focusing the user's attention on the appropriate parts and sequences. [17]. A study of instructions for constructing origami figures suggests that instructions that show each step were more effective than instructions that only showed the beginning state of construction, and what the final product would look like [13]. In the former case each step is depicted, while in the latter case the user must infer the intermediate steps. Although Black et al. [7] argue that manuals that allow users to infer steps provide an opportunity for deeper comprehension of the material in the manual, the task of assembling an object once is different from learning how to operate or troubleshoot a system. The end goal of the manual for assembly sequences is to perform effective and efficient assembly, not to acquire a deeper understanding of the system. Thus, the goal is to augment the performance of the user in a one-time task.

To understand the type of information conveyed in instructions for object assembly, Bieger & Glock [5, 6] created a taxonomy of information categories used in picture-text instructions and subsequently tested performance on assembly tasks with variations of categories represented in pictures and text. Results indicated pictures were more effective for contextual information, whereas for spatial and operational information, there were either no differences between modality, or a speed-accuracy tradeoff. These findings were considered in Feiner and McKeown's COMET system [9] and Andre and Rist's WIP system [3].

Other approaches to aiding users in the assembly process have augmented the user or the object to be assembled. Tang et al. [18] describe a system which show's the user where to place the next part of a Duplo assembly. Antifakos et al. [4] describe a system which instruments furniture parts sensors and allows the computer to guide the user by monitoring the assembly process. These systems are concerned with novel ways of presenting an instruction plan to a user. However, our focus is not on novel methods of presentation, but on determining what makes a good set of instructions. The plans produced by our system will be able to be used by these other systems.

Figure 1. The TV stand used in our experiments.

EXPERIMENTAL METHOD: STAGES 1-3

For Stages 1-3, we used the same basic experimental setup. We describe the setup here and then present the experiment performed in each of these stages.

We chose a TV stand as the object to be assembled (see Figure 1). To participate in these experiments, participants could not have previously assembled this particular model of furniture.

Spatial Ability and Experience (SA/E)

Individual differences such as prior experience in assembling objects and spatial ability affect performance in tasks that require mental transformations. To assess the roles of these factors, participants completed a questionnaire about their prior experience with assembling or building objects, such as model airplanes, Legos, dollhouses, or other toys. Participants also completed 2 tests of spatial ability, the Vandenburg test of mental rotation and the Money Spatial Navigation Task, a 1-minute test that evaluates egocentric perspective transformations [12, 22]. We combined the scores to form a single measure, spatial ability and experience (SA/E), as they were highly correlated, $r(41) = .532$, $p<.01$. Participants had to perform below average and have less experience to be included in the low spatial ability/experience category, and above average to be categorized as high SA/E.

STAGE 1: PRODUCTION

Our goal for stage 1 was to obtain hand-generated instructions produced by participants with both high and low spatial ability and experience. We did this by having users assemble the TV Stand from an unlabeled photograph and then asking the users to produce instructions for a novice assembler.

---

[1] The experiments conducted in stages 1,2 and 3 and described in detail in this paper were performed before we built the algorithms presented in [Agrawala et al. 2003].

Participants

Forty-five Stanford University undergraduates participated for monetary compensation. The data of two participants were eliminated as they had participated more than once, leaving 43 participants.

Procedure

Participants were tested individually. Participants assembled the TV stand without assembly instructions, given only a picture of the completely assembled TV stand. Upon successful completion, participants wrote instructions for assembling the TV stand. They were told to write instructions using text, diagrams, or a combination so that a novice assembler could easily and efficiently assemble the TV stand. They were given two pieces of paper on which to write the instructions.

Results and Discussion

Based on their performance on the spatial ability tasks and experience measures, we categorized participants' scores using a median split, yielding 21 low and 22 high SA/E participants.

Assembly performance

All participants were able to assemble the TV stand without instructions, and their assembly times were recorded. Because of the high amount of variance between participants, the time data were non-normally distributed. Thus, in order to perform an analysis of variance, which assumes error variance is independently and normally distributed, logarithmic transformations were performed on the assembly times. Log assembly time for all participants was 2.2 minutes ($SD = .38$). There were significant differences between high and low SA/E in assembly time. Low SA/E participants took longer ($\log M = 2.5$, $SD = .29$) to assemble the TV stand than high SA/E ($\log M = 1.9$, $SD = .25$), $F(1,41) = 44$, $p < .01$. See Figure 2 for assembly times in minutes.

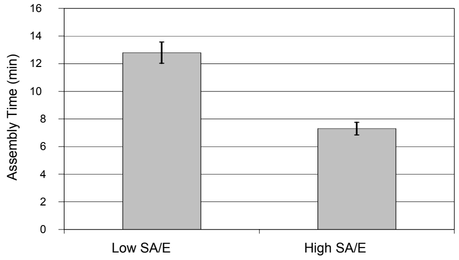

Figure 2. Assembly time in minutes by spatial ability and experience (SA/E). People with low SA/E took 12.76 (SE = .77) minutes to assemble, while people with high SA/E took 7.29 minutes (SE=.46) to assemble the stand.

Analysis of instructions

The TV Stand required a certain assembly order. For example, because the support board is held in place by pegs it is impossible to add the support board to the assembly after the left and right sides have been screwed in (See Figure 1 and Figure 7) steps 1 and 2 for visuals. Whenever instructions created by a participant violated such an ordering constraint, we counted the violation as an error. Errors in instructions produced by low SA/E contained more errors (86%) than those produced by high SA/E (12%) $t(1,41) = 5.9$, $p < .01$. 42/43 (98%) of participants included some type of visual depiction or diagram in their instructions. 26/42 (62%) of the diagrams represented information that was redundant with the text. The text often referred back to the diagrams. The diagrams drawn in the participants' instructions fell into 3 classes. First, people drew *structural* diagrams, that is, two or more parts in configured position. Structural diagrams were used to show a completed step or to demonstrate what an object should look like at a given point. Second, people drew *action* diagrams, that is, diagrams that represent one part joining another, demonstrating the necessary assembly actions. Figure 3 shows an example of a structural and an action diagram. Third, people drew *part menus* depicting the way parts look to differentiate two parts. Figure 4a shows an example of a part menu and Figure 4b an example of instructions produced by a participant with high SA/E.

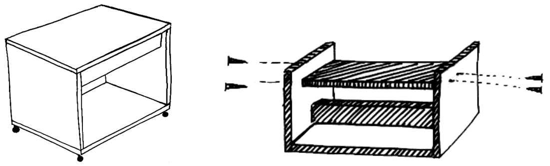

Figure 3. Examples of a structural diagram (left) and an action diagram (right). Notice that the action diagram spatially separates the parts that are to be added and uses guidelines to indicate the point of attachment.

	Parts	Structural	Action
Low SA/E	4.14 (2.6)	1.45 (1.4)	.64 (1.2)
High SA/E	2.2 (3)	.81 (1.1)	2.67 (1.9)

Table 1. Mean (Standard deviation) number of separate parts, structural, and action drawings per instruction set by low and high SA/E. Note that action diagrams contain structural information as well.

High SA/E produced more action drawings per instruction set than low SA/E $F(1,41) = 16.9$, $p < .01$. Conversely, low SA/E produced more structural diagrams than high SA/E though this difference was not significant due to high variance, $F(1,41) = 2.83$, $p = .1$. Action diagrams necessarily depict structure, so the majority of drawings produced by the high spatial participants depicted both action and structure. Low SA/E were more likely to include part menus compared to high SA/E, $F(1,41) = 5$, $p < .05$ (see Table 1 for means of these types of diagrams by SA/E). Diagrammatic visualizations are an integral part of instructions for visual and spatial tasks such as assembly. Participants of all experience levels agree that diagrammatic visualizations are important, as shown by the high percentage of participants in Experiment 1 that included them in the instructions (98%).

There were striking differences in the types of diagrams produced by high and low SA/E. High SA/E produced more action diagrams depicting the assembly process, used more 3-D visualizations and less text than low SA/E. High SA/E also made effective use of diagrammatic elements, such as guidelines and arrows to indicate placement or direction. Low SA/E had more errors in their instructions, used more structural diagrams

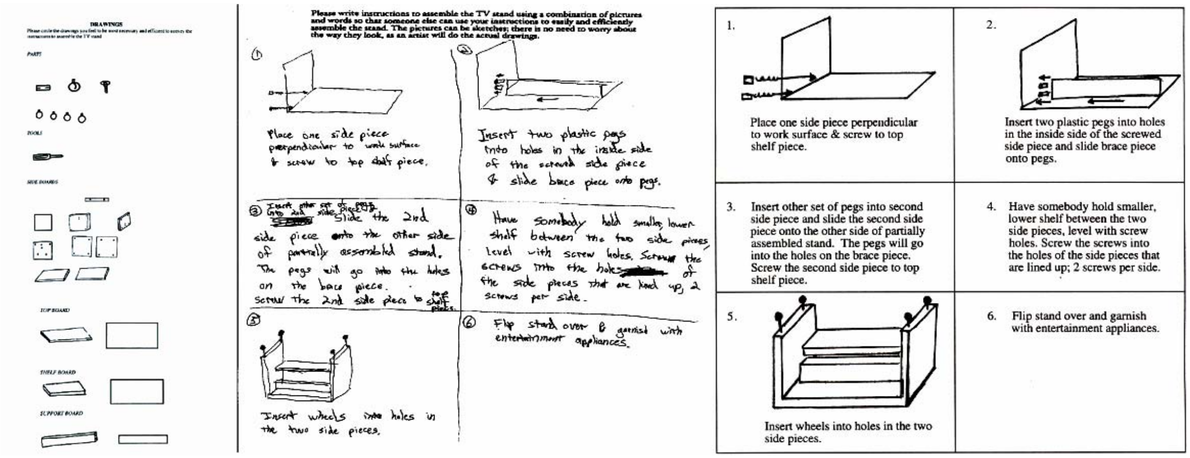

Figure 4a, 4b, and 4c. Examples of a parts menu (left), a set of instructions drawn by a user with high SA/E (center), and retyped and redrawn instructions used in Stage 2 (right).

and part menus, and more text to describe the assembly process. The results illustrate the importance of action diagrams in that they are used frequently in instruction by more experienced users. We hypothesize that action diagrams would augment the performance of low SA/E.

STAGE 2: PREFERENCES

Given the large differences in drawings produced by high and low SA/E participants, are there similar differences in their preferences? Here, we address this question by having a new set of participants rate a subset of the instructions produced in the first experiment.

Participants

Twenty-one Stanford University undergraduates participated in this experiment to fulfill a course requirement.

Procedures

As in Stage 1, participants completed an experience questionnaire and spatial ability tasks and based on a median split, were categorized into high and low SA/E. Participants then assembled the TV stand without instructions. After assembly, instead of producing instructions, participants rated a set of instructions produced in Experiment 1. Thirty-nine sets of instructions from the first experiment were rated. The other four sets were eliminated due to high similarity to other instructions in the sample, either using all text or containing similar diagrams. All written text was typed and sketches redrawn to equate clarity, legibility, and aesthetics (see Figure 4c). The order of instructions given to raters was randomized. Participants rated the quality and effectiveness of instructions from very poor (1) to excellent (7). They were also able to write comments explaining their ratings, specifically to describe the features of the sets of instructions they liked or disliked. In a post–task interview, they were asked to elaborate on features they liked or disliked. These comments were recorded.

Results and Discussion

By analyzing the variations in the instructions with the ratings and user feedback, we were able to extract characteristics of the instructions that were given high ratings compared to characteristics given low ratings. We did this in several iterations. First, we found the median rating (out of 7) across participants for each of the 39 instruction sets and categorized them as above or below the median. Second, in each of these groups, we deciphered the commonality between the instruction sets. For example, it was apparent that many of the low rated instructions did not have diagrams, or had part menus only. Similarly, in the high rated instructions, a commonality was the use of arrows and guidelines within diagrams to indicate action. Several passes were made through the instruction sets notating the commonalities amongst the instruction sets. For a feature to be considered, it must have been either a feature in common in the highest-rated (or lowest rated) instruction sets or based directly on the comments given by users. The relevant features, in order of importance, are presented in Table 2. The analysis was completed by 2 independent research-assistants. Results were compared and differences were resolved in discussion.

Amongst the participants, there was general agreement on the quality and the good and bad characteristics of the instruction sets (redrawn from Stage 1: Production). Ratings of the 21 participants correlated highly, $r = .97$, $p < .001$, and there were no differences in preferences of high and low SA/E participants. The top 5 instructions sets received a mean rating of 6, 5.8, 5.2, 5.1, and 5.0 out of 7, respectively. The redrawn instruction set that received the top rating (6) is shown in Figure 8a, and the instruction set received the second highest rating is shown in Figure 5. See Table 2 for a listing of the most salient features given high and low ratings

The high-rated characteristics are tested in Stage 3, as are characteristics that were preferred in varying degrees, such as the use of text and the use of exploded diagrams (one single diagram depicting all assembly actions). Stage 2 enabled us to understand what the users prefer. However, because their preference was reflective, meaning they think certain features would help them, it is important to have participants use instructions as they are assembling to validate their preferences.

STAGE 3: COMPREHENSION

The goal of stage 3 is to the further refine the findings from Experiment 1 & 2 of user's production and preference by examining how the assembly instructions are used in practice. The aim is to clarify the effectiveness of certain features, so as

High-rated features	Low-rated features
1. Action diagrams	1. No diagrams/parts only
2. Step-by-Step diagrams	2. Omitting steps
3. Clear order of steps	3. Order unspecified or impossible ordering
4. Views of relevant parts and attachments	4. Occluded views of parts and attachments
5. Use of diagrammatic elements to indicate actions (ex. arrows and guidelines)	5. Lengthy text

Table 2. Top five characteristics of high and low-rated instructions (in order of importance) from Stage 2: Preferences. See Figure 9 for the highest rated instruction set.

to create design principles that are based on production, preference, and actual comprehension. Stage 3 tests the comprehension and efficacy of instruction sets selected from Stage 1 and 2. These instruction sets were reconstructed in order to test different features of instructions, with the goal of extracting principles for the design of visual assembly instructions. A post-task interview was conducted to get feedback from the users.

Participants

Forty-four undergraduates from our institution participated in this experiment to fulfill a course requirement. They were randomly assigned one of four sets of instructions from which to assemble the TV stand.

Procedures

As in Stages 1 and 2, participants completed a questionnaire about their experience with assembling objects and two spatial ability tests, and based on a median analysis were categorized as high or low SA/E.

Results from Stage 2 gave good indications of features of instructions that users thought would be helpful to them. We were able to create 4 sets of instructions that reflected the preferences of users. These four sets of instructions were selected that included features that determined high ratings from Experiment 2. Each instruction set also included features that were preferred in varying degrees, such as use of explanatory text, presence of menu of parts, and integrated text and diagrams. Instruction set 1 had text and diagrams, describing and depicting the assembly process (See Figure 5a). Instruction set 2 used action diagrams only, with use of diagrammatic elements such as arrows and guidelines to indicated movement and attachment points. Instruction set 3 included action and structural diagrams, with a part menu (See Figure 5b), and instruction set 4 had a single "exploded-view" diagram with text describing the steps. Participants received 1 instruction set and were told to use the instructions provided to assemble the TV stand. As in Stage 1 & 2, participants were also able to use a picture of the completed TV stand. All participants were videotaped. After assembling the TV stand, they were queried as to the features of the instructions they found helpful or not. They were also shown the other sets and asked which features they thought would have been more or less helpful.

Results and Discussion

In addition to total assembly time and errors, we also coded how long they took for each step, and when they referred to the instructions. Log transformations were calculated on the assembly time data. An ANOVA was performed on the data, with between subject factors being SA/E and instruction condition.

As in Experiment 1 and 2, high SA/E assembled the TV stand faster (logM = 1.9, SD = .35) than low SA/E (logM = 2.2, SD = .32), F(1,41) = 12.8, p<.01. However, there were no differences in errors in assembly.

Importantly, the video data revealed that high SA/E consulted instructions an average of 7.1 (SE = 1.1) times and low SA/E consulted them 8.9 times (SE = 1.3), p >.1. This difference is not statistically significant, indicating both groups used the instructions. The videotapes also revealed that both high and low SA/E relied heavily on the picture of the completed stand. We noted each time participants looked at the picture of the completed stand. Low SA/E participants referred to the picture on the box an average of 9.13 time (SE = 1.4) whereas high SA/E referred to it significantly fewer times, an average or 3.4 times (SE = .84), F(1,42) = 12.3, p <.01. Further analysis revealed that the low SA/E participants were consulting the box for Step 1 and Step 3 of the assembly process. Step 1 is the beginning of the assembly, where participants have to orient themselves with the parts and perform the first action. This result indicates that it is important to make the orienting process easier, for example, by creating more realistic mappings of the diagrams in the instructions to the actual pieces (color, 3-D, shading). Step 3 is attaching either the shelf or the top shelf (depending on order), which requires deciphering which location to put the larger shelf. These results indicated that it is very important to clearly illustrate the pieces and points of attachment at each step, as occluding the perspective of the parts or the location of the attachment may hinder performance.

There were no differences in assembly time by instruction condition; however, responses in the post-assembly interview indicated that participants could explain features they liked and did not like. Overall, participants found extracting information

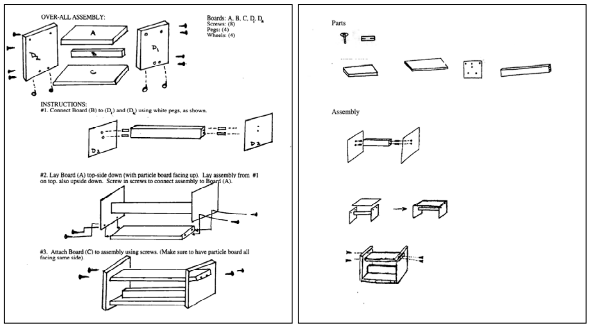

Figure 5a and 5b. Example of 2 instruction sets used in Stage 3: Comprehension. Instructions 5a were the second highest rated in Stage 2: Preference.

easier and more efficient from action diagrams than from text explanations. Several participants reported they never read the text explanations, that either it was too awkward to read as they were holding the stand up, or they felt the action diagrams were sufficient. Participants who received an exploded diagram were frustrated that order of assembly operations was not explicitly depicted, and because of the perspective, they couldn't see where one part was supposed to connect with another part. These results, particularly the feedback from the users, in addition to the findings from Stage 1 and 2 enabled us to create a set of design principles.

Design principles

In Stage 1, we analyzed the types of instructions produced by users of differing spatial ability and experience. In Stage 2, we analyzed users' preference of instructions and specific features of instructions produced in Stage 1. In Stage 3, we were able to validate the effective features of instructions by having participants use the instructions to help them assemble the TV Stand. Stages 1-3, respectively helped us to narrow and exact the features of effective visual instructions. The following 7 principles were important in each stage, in communicating to users the specific actions required to assemble the TV stand. Overall, the experiments provided guidelines that are directly relevant to assembly instructions and pertinent to other kinds of visualizations, particularly those that show processes over time.

- **Step-by-step, one diagram for each major step**: Each assembly action should be clearly depicted in a diagram and no steps in the assembly sequence should be omitted.
- **Clear and explicit order**: Assembly sequence should be made explicit by numbering the drawings for each step or placing them on the page in standard reading order.
- **Parts added in each step should be visible**: User should be able to know which part(s) is being added in each step.
- **Mode of attachment should be visible:** Where or how something is being attached should not be occluded by what is being attached.
- **Action diagrams rather than structural**: Action diagrams include structural information; depicting the action is imperative for assembly tasks.
- **Arrows and guidelines to indicate attachment:** Instructions should have consistent use of diagrammatic elements or semiotics to indicate motion or direction.
- **Avoid changing viewpoints**: The viewpoint of the object should stay as consistent as possible. If a new viewpoint must be chosen, it should be easy for the user to understand how to transform from the old viewpoint to the new one.
- **Show stable orientations:** It is important to show orientations of the object in a manner that is physically realizable.

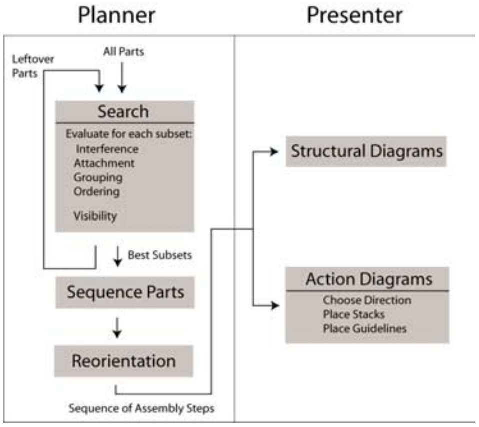

Figure 6. Diagram of the automated instruction design system.

STAGE 4: INSTANTIATION

We have implemented the design principles from the previous stage in an automated assembly instruction design system. The complete details of our implementation are presented in Agrawala et. al. [2]. We present a short summary of the implementation here.

Figure 6 shows a block diagram of our automated instruction design system. There are two modules, the *planner* and the *presenter*. The *planner* processes the input to determine a sequence of assembly steps required to assemble the object. The *presenter* takes the sequence of assembly steps and renders them as either structural or action diagrams.

System Input

The input consists of the geometry of each part assembled position, a default camera viewpoint and assembly orientation. The user may also optionally label parts as belonging to a semantically meaningful group (i.e. fasteners, wheels, etc…) and specify ordering constraints on the sequence in which parts are added to the assembly.

Planner

The planner is responsible for choosing the set of parts that should appear in each assembly step. It has three modules: *search*, *sequencing*, and *reorientation*. The search stage is responsible for selecting the best subset of parts to add to the assembly in each instruction step. It considers each subset of parts that have not yet been added to the assembly and rejects those subsets fail to meet any of a set of hard constraints.. The system then computes a score for the remaining subsets based on the visibility of each of it's parts.. The result of the search module is the subset that meets all the hard constraints and has maximal visibility.

Hard constraints consist of interference, attachment, grouping, and ordering. Interference and attachment check that it is physically possible to attach the subset to the assembly using geometric assembly techniques from robotics [15]. Grouping constraints check that parts belonging to the same semantic group are added at the same time. Ordering constraints ensure that the chosen ordering obeys the user-specified sequence.

To compute a subset's visibility, we define Visibility(P,Q) as the percentage of P that is visible with respect to Q, where P and Q are sets of parts. We compute three kinds of visibility:

- Current parts visibility (CPV). Each part in the current subset should be visible with respect to the other parts in the subset.
- Previous parts visibility (PPV). Some portion of the parts attached in earlier steps should remain visible for context.
- Future parts visibility (FPV). We want to ensure that parts added in an earlier assembly step do not occlude parts added in a later assembly step.

We require that CPV, PPV, and FPV meet minimum values to ensure that the parts added in each sequence have good visibility. The best subset of parts is the one that maximizes the total visibility, CPV+PPV+FPV.

Once the search has concluded, we have a sequence of assembly steps that specify the parts to be attached in each step. Before the presenter sees the sequence, we enforce additional cognitive design principles that are not addressed by the search. The *sequencer* ensures only one major step appears in each diagram. The *reorientation* ensures that if a change in perspective is necessary, we choose to show a stable orientation.

Presenter

The presenter renders the sequence of assembly steps as a series of numbered structural diagrams or action diagrams. Each structural diagram presents all parts attached in the current step along with all the parts seen in earlier steps in their final assembled positions.

As noted in Stage 2, action diagrams are easier to follow than structural diagrams. To generate an action diagram for an assembly step, we choose a direction that separates the parts in the current step from the earlier parts. We then set a separation distance for each part, and finally place the guidelines between the parts. Figure 7 shows a series of action diagrams automatically generated by our system to show how to assemble the TV stand.

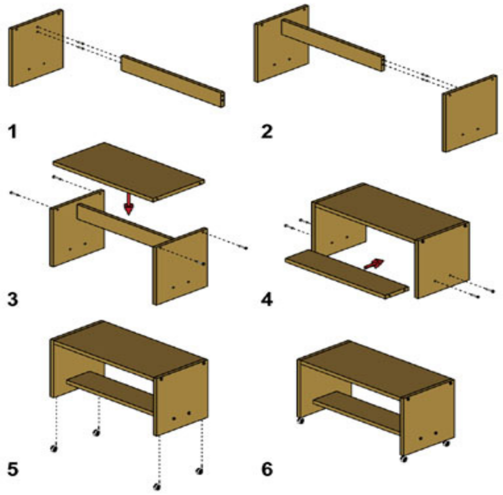

Figure 7. The instructions generated by our automated system.

STAGE 5: USER STUDY

The purpose of Stage 5 was to compare the instructions generated by the automated system described above to the top rated hand-drawn instructions from Experiment 3 and also the factory-provided instructions that came with the TV stand.

Participants

Thirty undergraduates from our institution participated in this experiment for monetary compensation. Participants were randomly assigned to one of three conditions receiving either hand-drawn, factory, or computer instructions. There were 10 participants in each condition.

Procedure

The pre-task procedures for this experiment were identical to those in Experiments 1-3. Participants completed an experience questionnaire and spatial ability tasks. Based on their performance on these tasks, they were categorized as high or low SA/E.

Participants were then given 1 of 3 sets of instructions. Set 1 was selected from Experiment 2, as it was given the highest ratings from participants in terms of its effectiveness. We will refer to Set 1 as the “hand-drawn” instructions. (see Figure 8a). Set 2 were the actual instructions that accompanied the TV stand, we will refer to these as the “factory” instructions (see Figure 8b). Set 3 were instructions generated from our automated system described above (see Figure 7). Participants were told to use the instructions to assemble the TV stand as quickly as possible (see Figure 9). They were not given the picture of the completed TV stand.

Results and Discussion

In Stages 1-3, we found that users relied heavily on the picture of the completed TV stand to assemble it. For example, they often used the picture more than the instructions in Experiment 3. In order to truly compare the effectiveness of the automated instructions with the other instruction sets, we did not provide the participants a picture of the TV stand, nor were participants told anything about what they were assembling.

A multivariate ANOVA was performed on the data with factors being instruction set (computer, factory, or hand-drawn) and SA/E. The dependent measures were the log transformations of assembly times and the number of errors made in assembly. The ANOVA revealed a significant main effect of condition, $F(2, 24) = 3.95$, $p<.05$. An LSD post-hoc test revealed that participants in the computer condition outperformed participants in both the factory and hand-drawn condition. See Figure 10 for assembly times comparison. As in Experiment 1, High SA/E assembled the TV stand faster than low SA/E $F(1,24)= 4.74$, $p<.05$.

In addition, participants in computer instructions condition made fewer errors ($M= .5$, $SD = .71$) than in the factory ($M = .6$, $SD = .97$) and hand-drawn condition ($M = 1.6$, $SD = 1.4$) $F(2, 24) = 3.795$, $p<.05$. In a post-assembly questionnaire, participants in the computer generated instruction condition rated their instructions as less confusing than participants in the factory or hand-drawn condition. Participants in the computer instruction condition also rated the task as less challenging than participants in the factory or hand-drawn condition. The results of the user

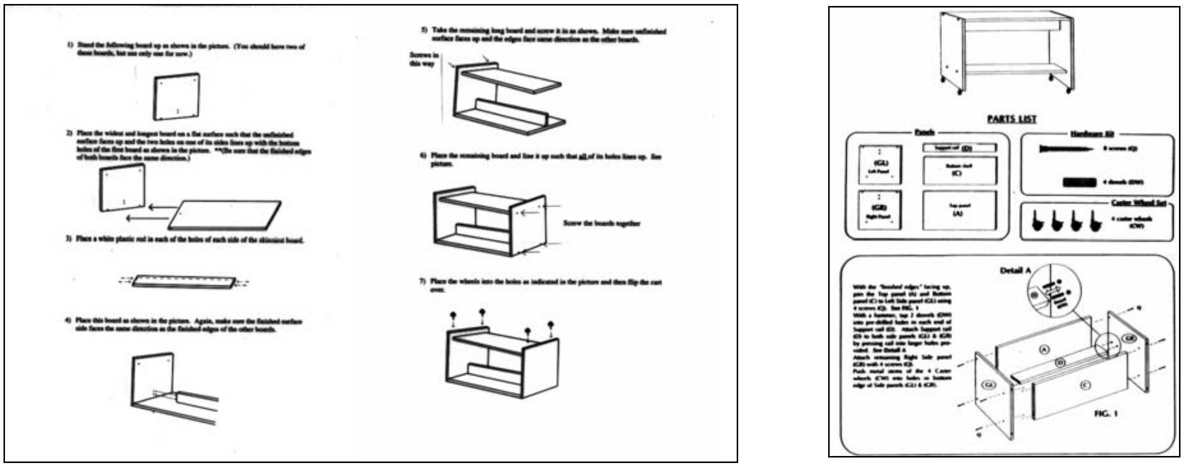

Figure 8a, 8bThe retyped, hand-drawn instructions (left) used in Stage 5: User Study. These were compared with the computer generated instructions shown in Figure 7 and the factory provided instructions (right).

Figure 9. A participant in Stage 5: Usability assembling the TV stand, referring to instructions generated by our automated system.

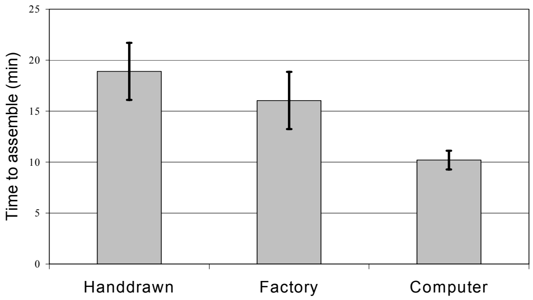

Figure 10. Time to assemble TV stand by instruction set. Instructions generated by automated system (M = 10.2, SE = .929) outperformed the top rated instructions (M =18.9, SE = 2.81) from Experiment 1 and the factory made instructions that accompanied the TV stand (M = 16.04, SE = 2.9).

study indicate that our system successfully instantiates the design principles revealed in Stages 1-3. Performance using our computer-generated instructions was significantly better, in both time and error rate, indicating that the instructions generated using the cognitive design principles aided users of both high and low SA/E in performing the task. Our results indicate that the computer-generated instructions informed by our cognitive design principles significantly reduce assembly time an average of 35% and error by 50%.

CONCLUSION

In education, psychology, architecture, and biology, visualizations are used to augment cognition and can be a useful instructional tool. Proper design of instructional visualizations can reduce the time taken and errors made in performing a task. To understand how to design effective visualizations we have identified and validated cognitive design principles for assembly instructions by investigating the production, preference and use of instruction in exploratory studies.

We developed procedures to reveal cognitive principles of effective instructions and implemented those principles into algorithms for generating individualized visual instructions on demand. We validated these principles by comparing the performance of users with our generated instructions to hand-drawn or those provided by the factory. The outcome of this final stage indicates the effectiveness of the instructions generated by our automated system, which thus reflect the approach to identifying the design principles. The design principles extracted from the current project, we believe, can be applied to the design of visualizations in similar domains.

Assembly instructions are representative of a large class of visual instructions including the assembly or operation of any complex system that consists of parts in a configuration. The cognitive design principles revealed by our study have generality beyond our test case. We look forward to extending this approach to other systems and domains. Developing instructional visualizations in educational domains, such as scientific visualizations like the human respiratory system, would provide a rich area for improvement in visualization design. Our work thus far has dealt with assemblies that only require simple linear motions, however we are interested in investigating how to effectively present more complicated assemblies to users that may require nonlinear motions. Perhaps for assemblies that require more complicated motions, it may be helpful to use animation as opposed to diagrams. The top-down and user-centered approach described in this paper provides what we believe to be a promising direction in visual interface design.

REFERENCES

1. Agrawala, M., and Stolte, C. Rendering effective route maps: improving usability through generalization. *Proc. Siggraph 2001*, ACM Press (2001), 241-249.

2. Agrawala, M., Phan, D., Heiser, J., Haymaker, J., Klingner, J., Hanrahan, P., and Tversky, B. Designing Effective Step-By-Step Assembly Instructions. *Proc. Siggraph 2003*, ACM Transactions on Graphics (2003), 828-837.
3. Andre, E. and Rist T. (1993). The design of illustrated documents as a planning task. In Maybury, Mark T. (ed.) Intelligent Multimedia Interfaces. MIT Press. 117-138.
4. Antifakos, S., Michahelles, F., and Schiele, B. Proactive Instructions for Furniture Assembly. *Proc. Ubicomp 2002.*
5. Bieger, G.R. and Glock, M.D.(1984-1985). The information content of picture-text instructions. *Journal of Experimental Education,* 53, 68-76.
6. Bieger, G.R. and Glock, M.D. (1986). Comprehending spatial and contextual information in picture-text instructions. *Journal of Experimental Education,* 54, 181-188.
7. Black, J., Carroll, J., and McGuigan, S. What Kind of Minimal Instruction Manual is the Most Effective. *Proc. CHI 1987*, pp. 159-162.
8. Denis, M. The description of routes: A cognitive approach to the production of spatial discourse. *Cahier de Psychologie Cognitive*, 16(4): 409-458, 1997.
9. Feiner, S. and McKeown, K.R. (1990). Generating coordinated multimedia explanations. In Proceedings of the sixth conference on AI applications, pages 290-296.
10. Mackinlay, J. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics 5*, 2, 110-141. 1986.
11. Mijksenaar, P. & Westendorp, P. (1999). *Open here: the art of instructional design.* London: Thames and Hudson.
12. Money, J., and Alexander, D. (1966). Turner's syndrome: further demonstration of the presence of specific cognitional deficiencies. Journal *of Medical Genetics*, 3, 47-48.
13. Novick, L. and Morse, D.L. (2000). Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition*, 28(7), 1242-1256.
14. Richards, C (2000). *Getting the picture: Diagram design and the information revolution*. Information Design Journal, 9(2).
15. Romney, B., Godard, C. Goldwasser, M., and Ramkumar, G. An efficient system for geometric assembly sequence generation. *Proc. ASME international Computers in Engineering Conference*, 699-712, 1995.
16. Seligmann, D. D., and Feiner, S. Automated generation of intent-based 3D illustrations. *Proc. SIGGRAPH 91*, 123-132.
17. Szlichcinski, C. Factors affecting the comprehension of pictographic instructions. *Information Design: The Design and Evaluation of Signs and Printed Material*. John Wiley & Sons Ltd. 1984.
18. Tang, A., Owen, C., Biocca, F., and Mou., W. Comparative Effectiveness of Augmented Reality in Object Assembly. *Proc. CHI 2003*, pp. 73-80.
19. Tversky, B., & Lee, P. U. How space structures language. *In C. Freksa, C. Habel, & K. F. Wender (Eds.), Spatial Cognition: An interdisciplinary approach to representation and processing of spatial knowledge*. 157-175. Berlin: Springer-Verlag, 1998.
20. Tversky, B., & Lee, P. U. Pictorial and verbal tools for conveying routes. *In Freksa, C., & Mark, D. M. (Eds.). Spatial information theory: cognitive and computational foundations of geographic information science*. 51-64. Berlin: Springer. 1999.
21. Tversky, B, Zacks, J., Lee, P. U., & Heiser, J. Lines, blobs, crosses, and arrows: Diagrammatic communication with schematic figures. *In M. Anderson, P. Cheng, and V. Haarslev (Eds). Theory and application of diagrams*. 221-230. Berlin: Springer, 2000.
22. Vandenburg, S.G. and Kuse, A.R. (1978). Mental rotations. A group test of three-dimensional spatial visualization. *Perceptual Motor Skills*, 47, 599-604.