# Delivering Instructions for Inherently-3D Construction Tasks: Lessons and Questions for Universal Accessibility

**Laura Leventhal**
Computer Science Department
Bowling Green State University

Bowling Green, OH 43403
+1 419 372 2765
leventha@cs.bgsu.edu

**Abstract**
*The notion that usability follows from the three dimensions, user, system and task, is not new. Clearly any model of effective user-centered design includes some focus on task; our work is an attempt to refine this notion. In this paper, I discuss a class of problems that we have dubbed inherently-3D construction tasks. These tasks should lend themselves to dual presentations, particularly those that include user-controlled 3D models or realistic videos. The results of several studies are described; the results indicate that dual presentations are effective at presenting instructions for inherently-3D construction tasks. When embedded in dual presentations, 3D models may be effective in presenting instructions for these types of tasks as well, so long as users make use of the models. A number of discussion questions are posited as to how this work would transfer to a wider population of users than was included in the studies that are reported.*

**Keywords**
*user-centered design, dual presentations, multiple representations, user-controlled 3D models, inherently-3D construction tasks, virtual reality*

## OVERVIEW

This workshop, "Universal Accessibility to Ubiquitous Computing." will focus on issues of accessibility to under-served groups, such as the elderly. Much of the discussion will likely focus on the relationship of user-centered design to issues of user characteristics, with system (user interface) characteristics. My paper will take a different track and focus on the interaction of task with user interface. The notion that usability follows from these three dimensions, user, system and task, is not new. Eason [2] makes the case that usability is defined in terms of system, task and user characteristics. Clearly any model of effective user-centered design includes some focus on task; our work is an attempt to refine this notion. I hope to raise some questions of how task characteristics might also play into design decisions for universal access.

At Bowling Green State University, we have been studying user issues for different presentation styles and for a variety of tasks. Our current focus is on the effectiveness of different combinations of informational representations for a category of tasks, which we call *inherently-3D construction task*s. Paivio's notion of dual coding is the theoretical basis for much of our work. [8] This theory suggests that visual and verbal presentations may be used effectively in combination. To date we have completed a number of studies of multiple informational representations that include visual and verbal elements.

In my paper, I will first discuss the characteristics of the systems and tasks that we have been studying. I will then briefly overview the theory of dual coding. Next I will review some of our studies and results with the idea of raising discussion questions of how this work might be extended to support universal accessibility.

## SYSTEM CHARACTERISTICS: MULTIPLE REPRESENTATIONS

Clearly a discussion of universal accessibility should include discussion of the World Wide Web, since for many users, the Web is a major part of their experience. From the perspective of our research, one of the fascinating characteristics of the World Wide Web is its multidimensional richness. A wide variety of media representations have become accessible via World Wide Web browsers. These newer representations have the potential to greatly expand the types of problems and solutions that are deliverable on the Web. Making the choice of media representations and their combinations is often non-trivial; recent advances in computer hardware and software have led to the development of a number of representations including: text, static images, streaming audio/video, computer animation and virtual reality (VR).

Many of the information representations now widely used on the Web are simply adaptations of existing media formats. For example, text and animation has been used for many purposes. There are also a small number of new technologies, such as 3D modeling and VR, which have truly expanded the types of information representations that are available via the Web. However, to date most of the widespread use of 3D graphics on the Web has been somewhat trivial and on the order of animated logos. There appears to be a general lack of understanding among developers of how to use and integrate 3D graphics effectively for various problem domains. In other words, it is unclear as to what types of tasks and users are best served by presentations that incorporate 3D graphics with other representational forms. [4] [7] has made a similar point about VR, stating 'virtual reality (VR) has generated much excitement but little formal proof that it is useful. Because VR interfaces are difficult and expensive to build, the computer graphics community needs to be able to predict which applications will benefit from VR'.

## TASK DOMAIN: INSTRUCTIONS FOR INHERENTLY 3D CONSTRUCTION TASKS

There is a class of problems that we believe would benefit from the use of multiple representation presentations that included some 3D representations. These problems involve the delivery of instructions to build a real-world object when the real-world object has some "inherently-3D" features." For example, we believe that the task *assembling a model airplane* is an inherently-3D construction task; delivering instructions for this task would fall under the class of problems that we are considering. Under our definition, inherently-3D construction tasks have the following characteristics:

1) The goal is to build a real 3D object. 2) Construction of the task requires a series of steps. 3) The object is asymmetric or changes symmetry during construction. In the case of a model airplane, there may be a passenger exit on only one side of the plane. See [12] for a more detailed description of inherently-3D construction tasks...Instructions for inherently-3D construction tasks have traditionally been delivered in a paper form; these instructions typically use combinations of text and still pictures to present the instructions. When instructions for these types of tasks are ported to the Web, they usually have retained their relatively static look and feel, incorporating text, still pictures and possibly a movie. However, in considering the inherently-3D properties of the real-world object in these types of construction tasks, an important question would seem to be what combinations of representations would be effective?

To investigate this, we have chosen the construction of origami (paper folding) objects as our inherently-3D construction task. Creating an origami object is similar to assembling a model airplane in several ways. Most origami tasks have a number of steps. While assembling an origami object, it is often useful to view it from a variety of perspectives.

The paper folding task creates an artifact in the real world. Thus, making an origami object is representative of a very broad class of problems in which a user is given step-by-step instructions and is expected to build something.

In addition there is a complexity spectrum for origami objects, ranging from simple, basically 2D objects with only a few folds, such as a paper hat, to highly 3D objects such as an origami box or shrimp. This complexity spectrum provides us with a framework to investigate how the 3D characteristics of the real object are related to the 3D characteristics that are conveyed in the instructions.

Additionally, in terms of a benchmark task for usability research, origami has several useful characteristics. Paper folding is a task which is familiar to most people; most people have made a paper hat. However, many people do not have explicit experience with traditional origami and thus they would need to follow detailed instructions to actually create an origami object. For most people creating an origami object is self-motivating;

we find that subjects in our studies are generally compelled to finish folding, because they enjoy the task of building the origami object.

Finally, developing presentations to deliver instructions to construct an origami object is a relevant task for the Web. Traditionally, the instructions for foldings are delivered onpaper, using text and still pictures. eg. [3] However, numerous Web sites and commercial multimedia products have recently become available; these sites and products convey the directions for origami objects with text, pictures, video and sound. [1]

## THEORETICAL BACKGROUND. DUAL CODING

Allan Paivio first proposed the theory of dual coding in 1971, and then with modifications and extensions brought the theory to its most current, published form in 1986. The most general assumption of the theory is that cognition of language and non-verbal objects/events are each handled by separate, specialized subsystems (i.e. the verbal system and the imagery system). 'Separate' means that the two subsystems are structurally and functionally distinct. Functionally, they are independent in that either system can be

active without the other or both can be active in parallel. They are also functionally interconnected so that activity in one system can initiate activity in the other Examples illustrating the functioning of dual coding include: reiterating 'E equals mc2' without having a mental image of what the formula means (verbal on/imagery off), recognition of a face but inability to remember the name (verbal off/imagery on), and the fearful reaction generated by just the mention of the word snake (verbal activating imagery). [8]
Dual coding theory appears to have application to the issues that combined media present to Web users, and suggests that visual and verbal presentations may be used effectively in combination. A number of studies of the dual coding hypothesis and variations, outside of the Web arena have found support for the theory [5, 6]

## OUR RESEARCH

In our recent work, we have studied the effectiveness of different combinations of representations for presenting directions for paper-folding tasks. In general, we have found that presentations that include both visual and textual components are more effective than presentations which are either strictly textual or strictly visual. In particular, we have found that dual presentations that incorporate user-controlled 3D models and animations in the forms of simple virtual reality (VR) or realistic video can be very effective in the delivery of this type of instruction set.. In all of our studies, the subjects were college students, enrolled in classes at Bowling Green State University. In this section, I review some our more important results.

### Constructing an origami whale: Varying the combinations of representations

We conducted several studies in which the task was to construct an origami whale. The completed whale consisted of 25 folds and the instructions were given in a series of 12 steps. Steps one through five were essentially 2D in the sense that they were folds or fold-unfold combinations on a flat piece of paper. In steps six through twelve, the folds were 3D, in the sense that they were building a 3D figure. According to our definition, this paper folding task was inherently-3D, although of moderate complexity.
Subjects received two training units: one on paper folding and one our presentation. Our instructions for folding the whale were presented in a variety of combinations of representations, including text, still pictures and user-controlled 3D models. In all, there were seven combinations of representations, including text (verbal presentation), still pictures (visual presentation), 3D models (visual presentation), text plus still pictures (dual presentation), text plus 3D models (dual presentation), still pictures plus 3D models (visual presentation) , text plus still pictures plus 3D models (dual presentation).
The 3D models were constructed with the Virtual Reality Modeling Language (VRML) and Java. VRML is the *de facto* standard development language for displaying VRs on the Web.. Contemporary Web browsers, such as Netscape, have plug-in features which allow users to view VRML applications. Slider bar controls allowed subjects to manipulate the 3D model.
We measured the accuracy of subjects folding the whales in terms of correct and error folds. To further understand how subjects used the presentations, we videotaped the subjects during the task. O*ur* results provided strong support for dual coding theory. The dual presentations, regardless of the specific dual elements, led subjects to make the most correct whales based on the number of correct and error folds.
We were puzzled in some respects by our outcome because we had predicted that the dual presentations which included a user-controlled 3D model would be more effective than dual presentations with still pictures. In analyzing our qualitative data, we made a number of observations:

• Subjects who had access to the user-controlled 3D models did not necessarily manipulate the models. Rather they used them as still pictures during the simple folds or as an animated cartoon of the folding step during the more complex folds.
• The type of information conveyed by the two visuals appear to be different. The still pictures show the folds, frozen in time while the 3D model shows the folds dynamically. Both the still pictures and the 3D model are abstract presentations, but the 3D model has the capability to 'run' and to be viewed from many different angles.
• When the folds became more three-dimensional, at Step 6, subjects tended to go to the visual presentation first and use the text as confirmation.
• The 3D model is an abstract rendition, in the sense that there were no realistic hands in the presentation. The model does not give orientation information and is free-standing. Without orientation, some steps may have been difficult to interpret.

In a follow-up study, we added explicit fold lines to original 3D models (as opposed to dynamic fold lines). The performance of subjects on this subsequent study were telling.

The dual presentations with the 3D models gave better results than those with the still pictures, suggesting that subjects were able to use the 3D models as both still pictures and more dynamic presentations. However we found in this follow-up that subjects still were limited in their use of the 3D models as user-controlled models and were more comfortable using them as animated cartoons.

From these studies, we concluded that the dual presentations were superior to the non-dual presentations. When the inherent-3D complexity of the task increased, subjects who turned to the 3D models, even to run them as animations increased their chances of success over those who did not use the models. [4,9]

**Constructing an origami whale: Changing the interface**

Our next study addressed an obvious question: did users not manipulate our 3D models because they did not understand how to use the manipulation tools, in spite of training.

In this study, the task was again to fold the origami whale. Subjects saw the dual presentation of text, still images and the user-controlled 3D model. Half of the subjects saw the slider bar tool set that we had used previously. The other half used the native interface of the CosmoPlayer TM plugin for VRML. A Java Applet controlled the animation. We found that subjects in either group did equally well in constructing the whale. Subjects in both groups made extensive use of animation. Subjects in either group who actually manipulated the models generally had greater success than those who did not. [11].

**Constructing an origami whale: 3D models or multiple perspectives**

In another series of studies, subjects again folded the origami whale of 25 folds, presented in 12 steps. In this study subjects were shown either a presentation with text, still pictures and a user-controlled 3D model with fold lines or a presentation with text, still pictures and animations of the folds. The second presentation included multiple perspectives of the folding for the two most complex of the folding steps, and thus approached the level of information from the first presentation.

We found that subjects were equally effective at creating the whale, regardless of the presentation that they saw. The results of this experiment have convinced us that if the folding task is not of great 3D complexity, a simpler model may suffice. However we observed qualitatively that for one of the complex folding steps, subjects who actually moved the user-controlled model had a much greater chance of success than either those who ran the model for the step as an animated cartoon or viewed the multiple perspectives. The challenge, we concluded was to remind subjects to use the information when they need it.[10, 12]

**Other studies of paper folding: Realistic videos**

We also have conducted two other studies of paper folding; in these final studies the presentations were hypertext but not Web-based. In these final studies, we investigated the usefulness of dual presentations. However instead of user-controlled 3D models, subjects were shown, under some conditions, realistic videos of the steps for folding the object. For the next study, subjects used a commercial product called “Origami: The Secret Life of Paper." [1] The package has several components related to origami; we focused exclusively on the instructional component. Within the instructional component, instructions are presented on a single screen, simultaneously in three windows, including a line-drawing window, QuickTime movie window, and scrollable text window.

Additionally, a small picture of the finished paper-fold and a navigational tool to move between fold steps are included in the instructional screen. The user receives all instructions within this screen.

The line-drawing window shows drawn images of the paper and its folds. In the QuickTime movie window, the user can play, pause, and rewind a realistic movie of hands folding each step of the paper-fold. The movie starts with the paper in the last configuration from the previous step. The scrollable text window contains background information and text instructions for each step.

For our task, subjects folded an origami crane. The crane consists of a total of 19 folds.

These folds are distributed across eight steps. Steps 1-4 are essentially 2D steps of creating folds on flat paper. Steps 5-8 involve 3D manipulations of the paper. We felt that the inherent-3D complexity of the task was likely similar to the whale of our previous studies. For this study, we simply videotaped our subjects.

Our results showed subjects followed similar patterns to our whale studies. For the flat, 2D folds at the beginning of the task, they used the scrollable text and line-drawings. For the more complex steps, they primarily relied on the video. However they used it in an interactive manner. They applied a “play, pause, rewind” strategy to complete the folds, segmenting the video into mini-steps. In most cases where the subjects applied this strategy, they would watch enough of the video to see how one fold was completed. Then they would pause the video and make the fold. Upon completing the fold, they would resume the video. If a fold was exceptionally difficult, the subjects would rewind the video and observe the fold again. One subject primarily used the video as confirmation that each step had been completed successfully. Like our previous results, these study suggests that subjects will try to use the more dynamic presentation, once the task itself is 3D. [4]

In our final study of paper-folding, subjects constructed an origami seal. The instructions for the seal consisted of sixteen steps and approximately 40 folds. Once again in this study, subjects saw a dual presentation. For the visual component, we varied a realistic video of hands folding the paper vs. a line-drawing animation. We found that with the realistic video, subjects completed more steps successfully. Also of interest, the subjects who had the line-drawing animation instead of the realistic video repeated almost twice as many steps as those who had the video. We felt that these results were especially interesting. The seal would appear, just on the basis of number of folds and number of hidden folds to have more inherent-3D complexity than the whale or crane. Perhaps the orienting information from the realistic video helped to support this higher level of complexity. [4]

## DISCUSSION AND SUMMARY

Eason [2] makes a clear case that the usability of any system is the collective outcome of task, system and user characteristics. In this paper, we have reviewed our research in the context of the interaction of task and system characteristics. We defined a class of problems that we termed *inherently-3D construction tasks*. Our previous research has indicated that when delivering instructions for tasks of this type, dual representations are effective. User-controlled 3D models and animations are examples of visual presentations that can be used effectively with text. Realistic video with text may be effective as well.

One of the interesting results from our work was that subjects often did not take full advantage of the visual presentation or used it in ways not anticipated. We now theorize that subjects may be sensitive to the extra overhead that is required of them to use 3D models or long realistic videos. When the step in the construction task has “enough” inherently-3D features subjects do use and benefit from the additional realism and functionality of the model or video. We speculate that as the overall 3D-complexity of the task increases, then subjects would increasingly benefit from the models and realistic presentations. We speculate that the whale and crane folding tasks overall were at the threshold or cross point between complexity and overhead. In our future research, we intend to explore this issue further by studying origami objects that have more hidden and inverted folds than the whale or crane. We are also currently in process of trying to link a cognitive measure with specific characteristics of the origami object as a way to quantify “inherently-3D.”

## UNIVERSAL ACCESSIBILITY DISCUSSION QUESTIONS

1. Our work indicates that multiple representations which include both visual and verbal components enhance one’s ability to follow instructions for an.inherently-3D construction task, particularly if one of the visual components is a user-controlled 3D model. Our work is based on a subject population of college students. Do our results transfer to senior citizens users? In particular, do user-controlled images help at all or are they yet another burden?

2. If user-controlled images hinder the performance of elderly users on this type of task, is the hinderance cognitive, social or just a lack of familiarity with the environment that students would not have, due to their exposure to other forms of manipulation (eg. video games)?

3. We found that under a number of circumstances, subjects did not use the “high tech” presentations in ways that they were intended. For example, many of our subjects simply ran the 3D models as animated cartoons, in spite of training to the contrary. Users who saw realistic video tended to segment them or even view them as stopped still pictures. Would elderly users be willing to utilize advanced technologies? What type and degree of training would be required?

4. Are some inherently-3D tasks too simple to justify the complexity of user-controlled models? At what point do these tasks become complex enough to justify 3D models and VR? Is this point different for senior citizen users?

5. Can the process of user-centered design be enhanced by focusing on the relationships of certain categories of problems and interface styles? Where does the elderly user fall in this?

## ACKNOWLEDGEMENTS

My colleagues in this work are Dr.s Julie Barnes and Guy Zimmerman. The work reported here has been conducted jointly with them.

## REFERENCES

1. Cassady & Greene. (1996) *Origami: The Secret Life of Paper* ISBN 1-56482-104-8

2. Eason, K. D. 1984, Towards the experimental study of usability, *Behaviour and Information Technology*, 3 (2), 133 – 143.

3. Honda. (1965) *The World of Origami*, Japan Publications Trading Co., Tokyo

4. Leventhal, L., Barnes, J., Zimmerman, G., Weihl, L. .(1999) Multiple Web Representations of Instructions on a Procedural Task: An Empirical Study . Computer Science Department. Bowling Green State University.99-MAY-01.

5. Mayer, R.E. & Anderson, R.B. (1991) Animations need narrations: An experimental test of a dual coding hypothesis. *Journal of Educational Psychology*, 83(4). 484-490.

6. Mayer, R.E. & Anderson, R.B. (1992) The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4). 444-452.

7. Pausch, R., Proffitt, D. and Williams, G. 1997, Quantifying immersion in virtual reality. In Proceedings of *SIGGRAPH 97* (Los Angeles, CA, August 3 – 8), Computer Graphics Proceedings, Annual Conference Series, 1997, (ACM Press, New York), 13 - 18..

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

9. Zimmerman, G., Barnes, J., Leventhal (2000) Delivering Instructions for Inherently-3D Construction Tasks: An Evaluation of Authoring Environments for Multimedia Presentations. Computer Science Department. Bowling Green State University. 00-MAY-01.

10. Zimmerman, G., Barnes, J., Leventhal, L. (2000) Delivering Instructions for Inherently-3D Construction Tasks: An Evaluation of Authoring Environments for Multimedia Presentations *ICS2000 International Computer Symposium*. Taipei, Taiwan. December 6 – December 8, 2000.

11. Zimmerman, G. Barnes, J., Leventhal, L. (2001) Evaluating The Effectiveness Of Web-Based Delivery Of Instructions For Inherently-3D Construction Tasks: Does The UI Matter? Computer Science Department. Bowling Green State University. 01-JAN-01

12. Zimmerman, G., Barnes, J., Leventhal, L. (2001) Building User-Controlled 3D Models and Animations for Inherently-3D Construction Tasks: Which Tool, Which Representation? *EHCI 2001: The Eighth IFIP Working Conference on Engineering for Human-Computer Interaction*. Toronto, Canada. May 11-14, 2001