ABSTRACT

WIEBE, ERIC N. Visualization of Three-Dimensional Form: Alternative Representations of Multiple 2-D Projections. (Under the direction of Sharolyn A. Converse.)

The increased use of rendering and animation tools in conjunction with 3D modeling software has heightened the need for careful evaluation of how 3D models are represented on the computer display. The experiment's hypothesis is that both rate in which projections of a rotating 3D object are presented and whether the object is rendered as a line drawing or shaded will effect the mental representation of the object. The experiment factorially crossed three levels of projection presentation rate with two levels of rendering (line drawing vs. shaded). All levels of both independent variables were between subjects. The subjects' score on a mental rotations test score was used as a covariant. The subjects each viewed 40 displays representing different rotating objects and identified the objects through a forced choice pair selection. RT and error rate were measured for each selection trial. Data on a total of 72 subjects were analyzed using the ANOVA procedure. The results of the analysis revealed a significant main effect of the rate of presentation variable on RT with the fastest rate showing significantly better performance. The results also showed a significantly lower error rate for line drawing versus shaded. No interaction was found between the two independent variables. The results indicate varying presentation rate can be an effective tool in allowing quicker interpretations of an object. It is also recommended that the display technique be carefully matched to the complexity of the object being displayed and the capabilities of the computer being used to display it.

VISUALIZATION OF THREE-DIMENSIONAL FORM: ALTERNATIVE REPRESENTATIONS OF MULTIPLE 2-D PROJECTIONS

by

ERIC N. WIEBE

A thesis-equivalent submitted to the Graduate
Faculty of North Carolina State University in
partial fulfillment of the requirements for the
Degree of Doctor of Philosophy

DEPARTMENT OF PSYCHOLOGY

Raleigh

1994

APPROVED BY:

Co-Chair of Advisory Committee

Co-Chair of Advisory Committee

ACKNOWLEDGEMENTS

I would like to extend my gratitude to the members of my graduate committee: Don Mershon, Vince Foote, Barry Beith and especially my advisor Shari Converse. Their advise both in the design of the experiment and the reporting of the results was invaluable. I would also like to thank my family — both immediate and extended — for the support they provided during work on this phase of my graduate program.

TABLE OF CONTENTS

Page

LIST OF TABLES AND FIGURES

Introduction

With the ever-improving price/performance ratio of computer equipment, the projections of the object from a changing viewpoint can be generated fast enough to give the sense of apparent motion. This dynamic projective display of three-dimensional objects has become affordable to a broad range of computer users. Besides users in industry, educators involved in teaching engineering and technical graphics realize the potential of these powerful display tools for assisting students in developing their mental imaging skills (McCuistion, 1991; Wiley, 1990).

One approach to teaching students how to perform this synthesis would be to start with a 3-D object, choose a stationary viewpoint, and rotate the object in small increments between the principle orthographic (perpendicular) views. This demonstration can not only be done with a real object, but also with a dynamic sequence of graphics on a computer. Generalizing the process, computer animation techniques can be used to help develop graphical techniques to describe how objects undergo all types of changes over time: in location and orientation and also in shape and size (Zsombor-Murray, 1990).

In addition to generating dynamic displays, computers are just as capable of producing static graphic displays similar to ones that have historically been used in teaching engineering and technical graphics. The question arises: are dynamic or static graphic presentation techniques more appropriate for teaching various visualization skills? In addition, does using computer rendering techniques to differentiate between specific faces of the object help in understanding its form? Questions such as these certainly do not have obvious or straightforward answers. The mental imaging of three-dimensional form does, though, strike directly at issues of considerable current interest to professionals and educators involved with engineering and technical graphics.

This experiment looks at some of the issues raised about the role of the current generation of computer graphics capabilities in the mental imaging of three-dimensional form. The first display technique of interest is parallel versus serial presentation of the same group of orthographic projections of an object. The fundamental difference in these two types of presentations is that the parallel format presents all of the projections on the display at the same time whereas the serial format presents the projections one at a time. Within the serial format there is also the variable of speed of presentation of the projections; does it make a difference if the serial sequence is presented at a rate that allows for the perception of apparent motion? The second display technique of interest is whether the application of gray shading to faces of the object's projections assist in the interpretation of 3-D forms. Though researchers such as Barfield (1990) have found no improvement in the interpretation of 3-D forms displayed with shading over those presented as black line drawings, others such as Sanford, et al. (1987) have. A mixed conclusion such as this, coupled with the popularity of computer rendering tools, make it worth exploring these results.

Theoretical Background

Experimental psychology in the areas of cognition and perception have a rich tradition in exploring issues related to the mental imaging of three-dimensional form. In addition, more recent research in this area has been directed specifically at the design of computer displays. Perceptual and cognitive principles are critical for understanding Human-Computer Interaction (HCI) and, of most interest here, the use of the computer to assist in the mental imaging of three-dimensional form (Allen, 1982; Haber & Wilkenson, 1982). Though no existing theories point towards a definitive approach to evaluating the presentation techniques of interest, many current theories in these research areas do impact on various elements of the problem. What follows is an overview of some of the theories having an impact on the research problem.

Imagery

Though psychologists, philosophers and others have attempted to answer questions pertaining to mental imagery over the years, a major resurgence of interest took place among cognitive psychologists in the early seventies. Shepard & Metzler (1971), instead of looking at the nature of abstract, logical thought, looked at a more analogical issue: namely, how does one mentally image and manipulate a set of connected cubes.

As the result of a series of experiments that he and his colleagues performed, Shepard (1978) came to a number of broad conclusions. Though it may be unwise to think of the brain process underlying a mental image as some sort of picture, there is a close, analogous relationship between a mental image and a perceived object. Secondly, subjects were able to formulate a mental image of an object even if it is not physically perceived. Furthermore, subjects used very similar mental processes on the image whether or not it resulted from perceiving a physical object.

These results are important for technical graphics educators and professionals. The way 3-D objects are formulated as mental images influences the way they are understood and represented graphically. In addition, mental solutions to graphics problems can be formulated even in the absence of physical images representing some or all of the elements of the problem. That is, more than just pure perception is stimulating mental imagery. There are also cognitive factors, including previous experiences, that play a role in mental processes related to imagery.

Internal Representation

Imagery theory has led to considerable debate as to the nature of the internal representation of spatial information. According to Kosslyn & Pomerantz (1977), this debate has two foci: first, whether images are a structurally distinct form of internal representation, the second being whether they are also a functionally distinct representation. The authors then try to define the assumptions made by the imagery accounts. Central to the assumptions is that once the image is formed, it is treated holistically and is compared to percepts in a template-like manner. This concept — one of just many pattern recognition theories — has direct bearing on how one might use projections generated on a computer display to support mental imagery of a similar form: "One often-cited use of imagery, after all, is as a model for anticipating the effects of physical manipulations. Thus it is plausible that the imagery representational system would have evolved to mimic the sorts of gradual transformations that occur when one physically manipulates objects." (Kosslyn & Pomerantz, 1977, p. 68). The image may be treated holistically once it has been formed, but what is the process by which this image is constructed? The authors contend that the most viable approach is not to work with building blocks at the most irreducible, propositional level, but at a more effective, functional level of emergent properties.

Pomerantz (1986) argues that attempting to measure every objective feature of the physical stimulus is an unproductive approach to understanding how one perceives form. Forms are multidimensional and hierarchical in nature and their component parts interact in ways to create *emergent features*. He notes how compellingly Gestalt psychologists demonstrated this principle. Terms used to describe emergent features: contour, closedness, shapes, three lines forming a corner, etc.; all have a ring of familiarity to those persons working with technical drawings of three-dimensional objects. Though most of the research in emergent features has focused on rather simple two-dimensional patterns, there may be identifiable emergent features that are crucial to the understanding three-dimensional form as it is represented on a two-dimensional computer screen (Treisman & Paterson, 1984).

Hochberg (1964) and Biederman (1987) also believe forms are perceived via their prototypical elements such as edges and corners. These prototypical elements —

whether they are edges and corners or other features — may also be the building blocks of the internal representations. For a person to perceive an object in motion, photo realistic images are not as important as images that contain the characteristic elements that make up its *canonical form*. That is, the canonical form becomes the internal structure into which one gathers the information from successive images and ultimately understand the nature of an object's three-dimensional form.

Retinal Patterns

Another approach to the concept of constructing a three-dimensional representation from a sequence of stimuli can be seen in the work of Gibson (1966). Rather than discussing features within an individual stimulus, the focus changes to patterns created through continual retinal stimulation. He concluded that these patterns, or *optic array*, contain enough information to fully comprehend three-dimensional form in the environment around us. An important point he makes is that normal perception is the visual world — our full comprehension of the environment around us — and not simply a series of visual fields or sensations. Each visual field, a discrete sampling that forms a retinal image should not be thought of as discrete, but as an overlapping mosaic. Comparisons among these fields isolate what is *invariant* during these transformations. These invariants, in turn, afford an understanding of the three-dimensional objects making up the visual world. One doesn't perceive a succession of retinal patterns but, rather, the three-dimensional form resulting from this scanning. This approach represents an even more holistic approach to three-dimensional form recognition. Not only are features not part of the vocabulary of image formation, but individual frames (or fields) are only part of a continuum and not to be processed independently.

A number of contemporary researchers are working on a computational model of perception that addresses the issue of interpreting changing retinal patterns. Marr (1982) considers that, from a historical perspective, Gibson came the closest to the level of computational theory. Gibson asked the critically important question of how does one perceive constancy in the world around us on the basis of continually changing sensations. One internally works with a representation of an object that constitutes its "real shape", not one based on a single retinal pattern.

It is at this point that Marr diverges from Gibson. Marr agreed that people are actively searching for physical invariants in patterns that add to our understanding of a three-dimensional form, but that this problem is exactly and precisely an information processing problem and that the detection of these invariants is vastly more complicated than Gibson has stated. Ullman (1979) goes even further than Marr in distancing the computational and Gibsonian approaches. Ullman felt Gibson's 'immediate perception' approach left no room for many of the fundamental tenets of the computational framework he was advancing. According to Ullman, the computational model can be viewed as a symbol-manipulating system. The computation it supports is the construction of useful descriptions of the visible environment.

There are also ways of looking at retinal stimulation in terms of frames but not necessarily with the use of features. Lappin, Doner, & Kottas (1980) showed that three-dimensional structure could be perceived by an individual in two frames of 512 random dots distributed over the surface of a transparent sphere. The authors concluded that it was the single projective transformation of this pattern of random dots that provided the necessary information for detecting the structure. Further experiments (Lappin, et al., 1983) continued to look at the issue of how an invariant structure could be perceived strictly from the projective transformations of patterns. These patterns may or may not contain what might be considered recognizable features. These theories give rise to the question what type of geometric or mathematical model best describes the changes in the retinal image resulting from the object in motion?

Geometric Models

Much of our thinking about the description of objects and space continues to be influenced by Newton's and Kant's view of three-dimensional Euclidean space (Eliot, 1987). It is a conception that gives rise to the notion that one can use an extrinsic measurement system (usually linear and two-dimensional) to directly measure the retinal image. A number of researchers have pointed out that when applied to the real-world case of geometry in motion, this Newtonian approach leads to many artificial complications (Johansson, 1975; Lappin, 1986). Working under the assumption the internal representation of a three-dimensional form is independent of any particular orientation, a geometric model which preserves the perceptual invariance of an object as it undergoes motion will be the most profitable approach. Given the temporal nature of an object undergoing motion, Eliot urges readers to consider non-Euclidean models which capture the spirit of Einstein's contention that timeless space does not exist.

Johansson (1975) advocates a complete break from the Euclidean model by proposing a nonmetric geometry based exclusively on relations rather than particular measurements. This geometry, *projective geometry*, is based on the concept that certain relations remain invariant under perspective transformation. Like Gibson, he believes these invariances in the changing retinal image are abstracted and transformed into percepts of rigid objects moving in three-dimensional space. Using a model titled *perceptual vector analysis*, Johansson has performed a number of experiments exploring phenomenon surrounding this concept. Though the model proposes a non-Euclidean method for processing the visual information, the resulting perception (from the internal representation) is as one would expect: constant Euclidean shapes in rigid motion in a three-dimensional world.

The computational model of Ullman (1979) also directly addresses the issue of objects in motion but from the standpoint of symbol processing rather than the spontaneous abstraction of three-dimensional form from changing retinal patterns. The first step is the matching of different images representing the same physical object by the correspondence process. Once this is done, one of two distinct interpretation processes is used to transform the images into a representation of a three-dimensional form. The first interpretation process, *Structure from Motion* (SfM), states that one has the capability to infer 3-D structure from a changing image when each static projection by itself contains no useful 3-D information. The fundamental problem underlying this interpretation process is the ambiguity that arises from the lack of a one-to-one correspondence between a given 3-D structure in motion and a given 2-D (retinal) image transformation. Constraints need to be applied to the image, a primary one being that the image transformations should be interpreted with a bias towards understanding it as a rigid object.

Building on the rigidity assumption, Ullman derives his *structure from motion theorem*: "Given three distinct orthographic views of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined." (Ullman, 1979, p. 148). Under this system, local features are resolved through rigid interpretations applied to a nuclei of elements using an approximation of orthographic projection. The polar component of the projection system is then applied globally to the object to fully resolve the 3-D form of the object. The second interpretation scheme proposed by Ullman, *Motion from Structure* (MfS), recognizes that often 3-D structure is perceived from single static frames. Under this scheme, known structure can be used to derive motion in space from a series of static frames. It follows that when static 3-D perception is present, both SfM and MfS operate simultaneously. It is not perfectly clear, however, the extent which the static and dynamic interpretations interact.

Most of the research reviewed above agrees that the mind makes use of a singular, canonical form of three-dimensional objects that is independent of any specific projective view of the object. This internal representation can be formed by the stimulus of multiple projective views of an object and can be used to compare with other projective views to determine the similarity or difference between the internal model and the stimulus. There is however, no agreement as to how this internal representation is formed or is used to evaluate novel forms. For some, the building blocks derive from recognizable geometric and topological features such as the shape of a face or whether the contours form a closed polygon. For others there is a much more direct route from stimuli to percept with no perceptual identification of individual features from any particular projective viewpoint. Still others state that though features in and of themselves are not critical, the comparison of retinal patterns (whether they be "features" or not) from projective transformations is central to evaluating three-dimensional form. A central issue still not directly addressed is the advantage of displaying multiple projective views in a serial or parallel format. The metaphor of the mind as an information processor could be useful in better understanding the interaction between the operator and these two types of displays.

The Mind as an Information Processor

The analogy of the human mind as an information processor affords some unique approaches to understanding the generation of internal images or representations. An approach to this information processing model is to look at the "bandwidth" of the mind to receive and process information (Baecker, 1987). Perception, as an active process, needs resources to process visual information. When these critical resources are limited, our ability to make use of visual information degrades. Though there is a point of diminishing return, if more resources are made available, performance will be enhanced. If no more resources are available, another way to increase performance would be to improve the quality of information being received so that less processing has to be done. Yet another way to improve performance is to be able to rely on "previous experience" to assist in the processing of the information. The efficiency with which this information can be retrieved from long-term memory will influence its usefulness in processing current information.

Kosslyn & Pomerantz (1977) contend that there is only finite processing capacity available for constructing and representing images. Functionally, this manifests itself in the amount of detail that is available about a form. They go further by stating that the mind has the ability to allocate resources to different spatial locations in a mental image. That is the center of an image may be in 'sharper focus' because more capacity is allocated for the construction of central details.

Others would argue that reaching certain input thresholds alters the processing of the visual information. Considerable processing can be done on retinal patterns with purely perceptual mechanisms according to Julesz (1975); these are mechanisms that are performed spontaneously without any cognitive processing stages that involve scrutiny. At the same time there is a distinct limit to how much pure perception can contribute to form recognition. Based on research results, Julesz contends that the purely perceptual pattern detection can only extract the simplest features of a form from an image and that for full comprehension of a form, higher level processing involving scrutiny is necessary.

Teichner & Mocharnuk (1979) note that the processing rate of stimuli in a display increases as the number of stimulus dimensions increases. Another important factor is the whether the stimuli consist of novel geometric forms and the number of different types of geometric forms in the display. Both of these can alter the strategy employed to process the stimuli. When looking at the interpretation of a 3-D object, Sanford, Barfield, & Foley (1987) and Yuille & Steiger (1982) both found that when figure complexity increased, the object was evaluated on a feature-by-feature basis rather than in parallel. This feature-based analysis is akin to symbol manipulation in the computational model of perception (Marr, 1982; Ullman, 1979)

Marr (1982) postulates that the retinal patterns generated through vision go through a number of processing stages in order to build up a final internal representation of an object. At the early stages, a *primal sketch* is formed through algorithms that make explicit important information within a retinal image. This information is made up of such primitives as edge segments, boundaries, curvilinear organization and the like. Later stages of processing are *2 1/2-D sketch* and finally *3-D model representation*. As in most of the previous models mentioned, this model takes into account both static and dynamic imagery. Dynamic imagery, in particular, has led to interesting models of how the final three-dimensional form is perceived from a stream of visual stimulation.

Paivio (1983) states that an image is a visual nested hierarchy based, in part, on levels of complexity. These features can be seen as components of the overall retinal pattern. If comparison of these components contributes significantly to the generation of the internal canonical representation, then techniques that enhance the viewer's ability to locate and organize these components of the overall retinal pattern should increase visualization performance. Both Hochberg (1964) and Paivio (1983) looked specifically at the role that long-term memory played in imagery and the importance of the likely differences between the ways visual and verbal information are manipulated. Pavio maintained that long-term memory contained two different but interconnected symbolic systems for coding and representing information. Imagery involves a memory code for processing spatially synchronous information whereas a verbal coding system processes in a sequential fashion. Hochberg postulated that this special structure, or schema, for processing spatial information is used both to build the canonical form of an object and to compare it with novel objects. This schema can also be used to anticipate what an object should look like from a new projection. When watching a dynamic sequence or scanning a group of projections, these comparisons are ongoing, either confirming or disproving one's anticipations.

In this review of the literature, a number of models inferring the internal mental structure of three-dimensional spatial information have been hypothesized. A description of any structure would be incomplete without an analysis of how the human mind receives, processes, and stores information within this structure. Information processing theory shows promise in shedding light on the dynamics of these proposed structures. Are there some graphic structures presented on a computer display which more closely parallel the internal structure used to understand three-dimensional form than others? Are there some methods of presenting these computer displays more attuned to dynamics of how the internal structures process this visual information? Are there ways of enhancing the computer displays through rendering techniques that enhance the processing of this visual information?

In this experiment, it is hypothesized that computer graphic display designs which enhance a person's ability to form an internal mental representation of 3-D objects will enhance performance in tasks involving graphic representations of these objects. More specifically, a person with a superior internal representation of an object should be able to more quickly and accurately choose the correct object from a forced choice pair of projections of two different objects. This experiment varied methods for representing the rotation of a 3-D object on a computer screen. There are two computer display design principles of interest: the temporal dimension in which projections of the rotation are displayed and the method of representing the edges and faces of the object. Of interest is which display design depicting the rotating object results in a superior internal representation. Because none of the theories discussed directly addresses the applied issues of interest, no one theory was chosen for hypothesis testing in this experiment. Each will be considered in relationship to the data obtained. Future experiments will be used to hypothesize about specific elements of theories supported by the results of this initial experiment.

Method

Subjects

Subjects consisted of 72 North Carolina State University staff and students. A majority of the subjects were students from the undergraduate psychology student pool. These subjects received partial course credit for participation in the experiment. No subjects had previously taken any technical or engineering graphics courses. In addition, the subjects were screened for normal visual acuity. Subjects were accepted into the experiment if they achieved at least 20/30 vision — with or without lens correction — on a visual acuity test administered with a telebinocular.

Design

The experiment employs a 3 x 2 full factorial design with three levels of Format (Dynamic/Sequential/Static) and two levels of Rendering (Line/Shaded) technique. Both independent variables were used between-subjects. Combined, the two independent variable dimensions represented six possible computer display design factors (see Table 1). Response time (RT) and accuracy (error rate) were the response measures. Previous studies by Vandenberg & Kuse (1978) have shown there is a small but consistent difference in the performance between males and females on spatial visualization tasks. For that reason, subjects were assigned to one of six groups (12 per group) counterbalancing for gender.

Apparatus

An Apple Macintosh IIcx with a 16 inch Viewsonic monitor was used to display the experimental stimuli. The monitor was set to only display gray scale colors. Subjects sat 45cm from the monitor with their eyes approximately level with the center of the screen. Ambient illumination was produced by an overhead fluorescent fixture and was held contest over the course of the trials. A standard Extended Macintosh keyboard was used with the computer.

Table 1. Matrix Representation of Experimental Design

	Format		
Rendering	Static Line	Sequential Line	Dynamic Line
	Static Shade	Sequential Shade	Dynamic Shade

The initial subject information, the instructions for responding to the stimuli, and the stimuli themselves were displayed and controlled through a script written in HyperCard 2.0. Timing for the stimuli was controlled by the script as was the recording of response selection and RT. These data along with the initial subject information were stored within the program and then written to a text file after subjects had completed viewing the 40 stimuli. An external program, QuickTime, was used in conjunction with HyperCard to drive the image sequences seen in the Dynamic and Sequential formats.

Experimental Stimuli

The present experiment looked at two different techniques, Format and Rendering, for displaying a three-dimensional object on a computer display. In all of the computer displays, the subjects looked at a three-dimensional object described through a series of orthographic projections. Orthographic projection is the historical method for depicting objects in technical and engineering graphics. Convergence, which adds a degree of realism in perspective projection, is not present in orthographics. Besides allowing for more simplified computation for generating the displays and ease in direct measurement of the object, both Braunstein (1986) and Ullman (1979) argue that convergence is not a significant factor in perceiving three-dimensional form.

Though researchers have shown that the detection of three-dimensional structure is possible in as few as two (Petersik, 1980) or four (Ullman, 1979) frames, a total of thirteen different frames were used in the study. The thirteen frames represented the object undergoing rotation in 15 degree increments about both the vertical and horizontal axes. This rotational sequence allowed for the display of all three primary orthographic views of the object (front, top, and right side) along with intermediate views at 15 degree intervals (see Figure 1). Though the rotational axes were chosen primarily for their significance to technical and engineering graphics, they also had perceptual significance. Green (1961) found rotation about the vertical axis allowed accuracy of perception of three-dimensional forms superior to any other rotational axes. Horizontial and skew axis orientations resulted in poorer performance, with no difference between them.

The stimulus information was presented in one of three levels of the Format variable. With the subject sitting at 45cm from the computer screen, the visual angle of a single projection of the 3-D object in an experimental display was 3.6 degrees. All thirteen projections shown on the screen subtended 29 degrees at this viewing distance.

The thirteen projections shown in parallel constituted the Static display; the content of the display did not change during the entire exposure interval. The organization of the Static display was the same as shown in Figure 1 and allowed for uninterrupted scanning of the projections as the object went through a 90 degree rotation about each axis. The Static display contained some resemblance to the concept of *Small Multiples* (Tufte, 1990) wherein a sequence of closely related graphic elements are arranged in a matrix. In this experiment, the two dimensions of the matrix were the rotational axes of the object. One of the principle features of the Small Multiples matrix is the ease in which nodes (individual projections in this case) are compared for changes. Individual projections were compared for changes in the shape or size of edges and faces on the object as it underwent rotation.

The second level of the Format variable (Sequential) showed the projections in a serial manner: only one projection of the object was shown on the computer display at a time (see Figure 2). The projections were shown at a slow enough rate (1 frame/sec) to preclude apparent movement in depth as it underwent rotation. Petersik (1980) concluded that temporal factors are much more critical than spatial factors in inducing the perception of apparent motion from a sequence of frames. Objects can rotate as much as 180 degrees between frames as long as the time between frames is kept under 300 msec. By having the interframe rate equal to 1sec, apparent motion was not induced in the Sequential format.

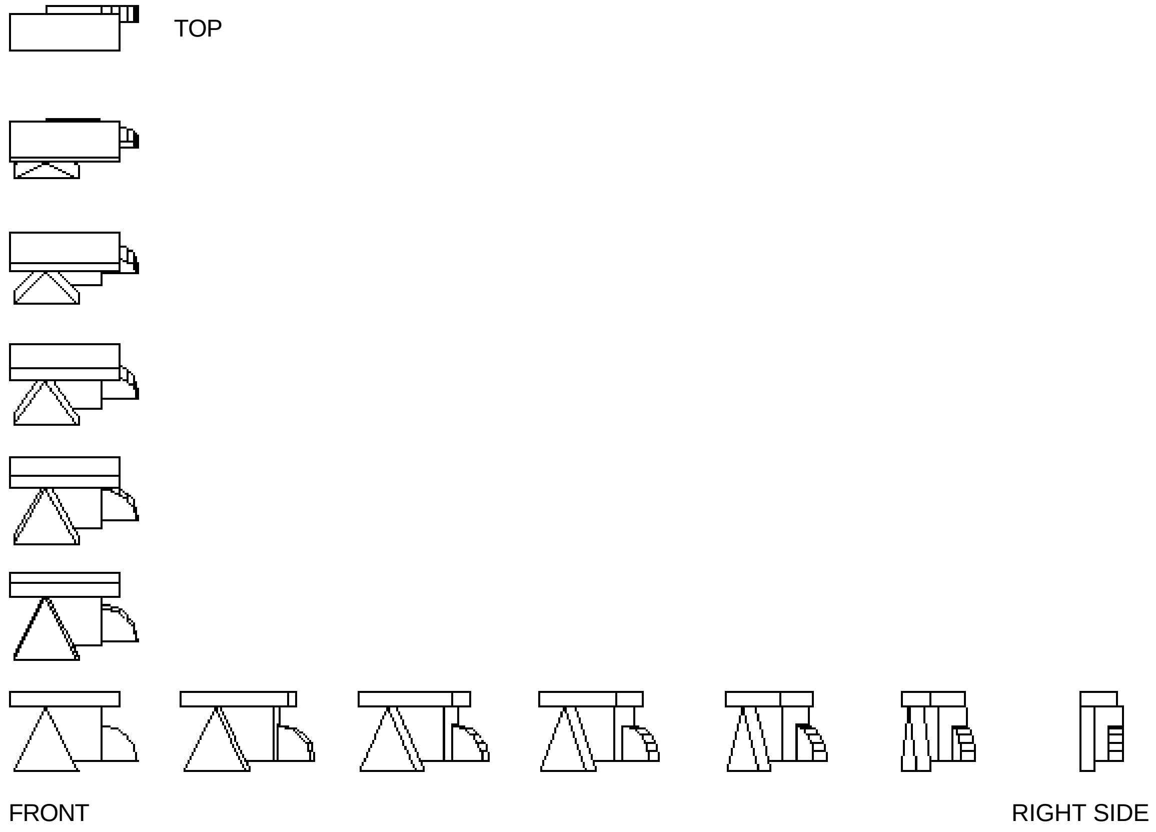


Figure 1. The Static/Line display design depicting all thirteen projections.

(Principle Views Labeled for Illustration Only)

The third level of the Format variable (Dynamic) was a serial presentation where the frames were shown at a fast enough rate (250msec/frame) to induce apparent motion. This presentation was akin to animations shown on computers or film.

Since all three levels of the Format variable were shown to subjects for the same total length of time, the Sequential level was the format determining the minimum display time. At 1 frame/sec for thirteen frames, 13 sec were needed to see all of the frames. The frames were presented in the Sequential format so the object underwent rotation from the right side view to front view and then to the top view. The Dynamic format, rotating at 4 times the rate of the Sequential, underwenta different sequence during the 13 sec. Presenting the frames at a rate that induced apparent motion, the object rotated back and forth between the right and front views three times, thereby producing a *rocking* motion (see Figure 3a). The object then proceeded to rock back and forth between the front and top views three times (see Figure 3b). The Static display simply stayed on the screen for the 13 sec.

Within each Format level, the display was varied by the second independent variable, Rendering. This second independent variable varied the representation of the edges and faces of the object. The first level of Rendering, called Line, represented the object's edges as black lines against a white background (see Figure 2). Any edges that would normally be obscured from view were not shown.

The other level of the Rendering variable was called Shade. The Shade rendering applied a contrasting gray shading to the faces on the object (see Figure 4). This rendering technique simulated the effects of an infinite light source projected onto the surface of the object.

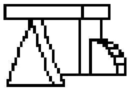


Figure 2. An single projection frame used in both the Sequential/Line and Dynamic/Line display design.

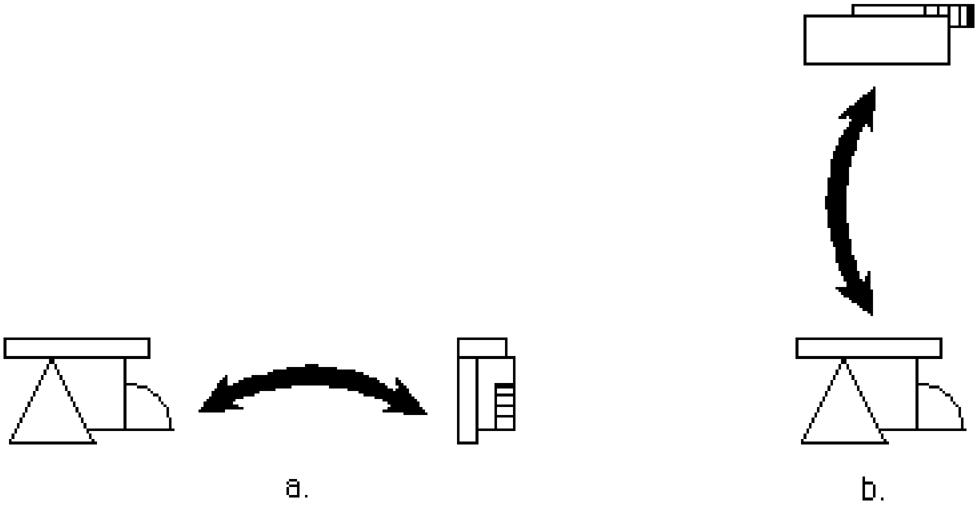


Figure 3. Projection frame sequence for the Dynamic/Line display design.

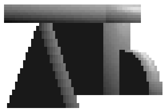


Figure 4. An single projection frame used in both the Sequential/Shaded and Dynamic/Shaded display design.

Procedure

When the subjects arrived at the assigned room for the experiment, they were seated and given a consent form describing the experiment. If they did not have any questions, they were asked to sign the form. The subjects were then given a visual acuity test with a Keystone telebinocular. If they scored at least 20/30 on the test, they were led into the next room and seated to the right of the computer.

Seated at the table, the subjects were given the Mental Rotations paper test (Vandenberg & Kuse, 1978), a 20 question paper test based on an experimental technique developed by Shepard & Metzler (1971) (see Appendix A). This test measured the subject's ability to mentally rotate three-dimensional objects and, it is hypothesized, provides a partial measure of the ability to do the type of task required in this experiment. The subject's score on the Mental Rotations test was used as a covariant in the data analysis. Subjects were given verbal instructions on taking the test and given a target time limit of 15 min to complete it. Once the paper test was over, they moved their seat over to the adjacent computer display.

The subjects were instructed to adjust the keyboard on the desktop to a comfortable position and to keep their fingers positioned over the arrow keys. On the computer, the subject's ID number, year in school, and his/her judgment (on a 5 point Likert scale) as to how good they were at visualizing 3D forms were entered into the computer database (see Appendix B). Once this information was entered, the subject read on the computer screen how to respond to the experimental stimuli (see Appendix C). These instructions indicated that subjects were to respond as quickly as possible but to try to correctly pick the object on at least 38 of the 40 displays they were going to see (95% accuracy rate). After completing the instructions, the subject pressed any arrow key and began the main part of the experiment.

The subject viewed an object represented by one of the six possible display designs for the specified time of 13 sec. After a screen blanking interval of 1 sec, a pair of different objects using the same rendering method was displayed side-by-side from the same projection (see Figure 5). One of the two objects was the same as the one they were just viewing. The wrong object varied from the correct one in topology and/or geometry. That is, both the number and configuration of faces on the object may be different along with the size and shape of them. Geometry and topology represent the two primary ways 3-D objects visually varied from each other. Whether the correct object was displayed in the right or left isometric pictorial in the forced choice display was randomized, appearing on the right 18 times and on the left 22 times. The visual angle for a single projection in the forced choice pair was 9.5 degrees.

The projection, an isometric pictorial, used to display the pair of objects was different than any of the projections seen in the experimental display designs. This projection was chosen for two reasons. First, using a different projection precluded comparisons between the stimulus object and the forced choice pair objects based on any specific projection seen in the experimental display. Second, the isometric projection — like the orthographic projections used in the experimental stimuli — represented a standard projection method likely to be encountered in a typical technical drawing.

The subjects responded to the forced choice pair by pressing either the right or left arrow key. Pressing the arrow key indicated whether the right or left object was the same as the one previously viewed. Immediately thereafter, they were asked whether they had high or low confidence in their response by choosing either the up or down arrow key.

Each group participated in one training block (10 objects) and one experimental block (30 objects). Both the training and experimental blocks used the same combination of Format and Rendering. Each group of subjects saw the same 40 objects in the same sequence.

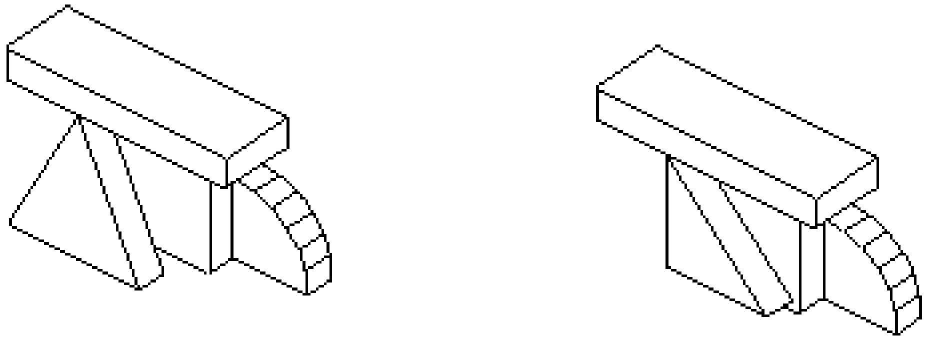


Figure 5. Forced choice pair of objects shown in isometric projection

After the practice block of ten objects, the subject was shown their average RT and number of correct responses out of the possible ten. The subjects continued with the experiment by pressing Return on the keyboard. After the experimental block of 30 objects, the subjects were again shown their average response time for all 40 objects and the number correct out of a possible 40. This time when they pressed Return, the initial title screen appeared on the monitor. The subjects were asked if they had any questions, given credit slips indicating they completed the experiment, and excused from the experiment.

After completing the experiment, subjects were excluded from the analysis if they failed to meet one of two minimum criteria. First, their score on the Mental Rotations test had to be at least 3 correct out of 20 questions (better than random guessing). If they failed to meet this baseline, it was assumed they lacked the necessary ability to perform the experimental task. Second, they were excluded if their error rate on the experimental task was more than 2 standard deviations above the mean of the rest of the subjects. This elevated error rate level was taken as an indication they failed to adhere to the stated strategy of achieving an 95% accuracy rate.

A total of 83 subjects participated in the experiment with 11 being excluded because they did not meet one of the two minimum criteria. Nine subjects were rejected because their Mental Rotations test scores fell below random guessing (less than 3 correct). Two other subjects were rejected because their error rate was more than 2.0 Standard Deviations higher than the mean rate of the first 72 subjects. The mean was 2.58 with a S.D. of 2.49. Those excluded had error rates of 8 or more wrong.

Results

Analysis

The error rate and RT data collected in this experiment were each subjected to individual two-way analyses of variance (ANOVA). The independent variables were Format and Rendering. For each subject, only the experimental block constituting the final thirty objects was analysed. The error rate was calculated as the total number of incorrect responses and RT as the mean RT for each trial block. These ANOVAs compared the error rate and RT for each combination of Format and Rendering shown in Table 1. In these analyses, the subject's score on the Mental Rotations test was used as a covariate. The score was recorded as the number correct out of 20.

Response Time

The analyses revealed a significant main effect for Format, $F(2,60) = 6.71, p < .0024$. The Format variable accounted for 15.5% of the variability in Response Time. The results of the ANOVA for RT is summarized in Table 2. No significant interaction was found between the two independent variables of Format and Render, $F(2, 60) = 0.25, p < .7778$; the Mental Rotations test score and Format, $F(2, 60) = 0.09, p < .9160$; the test score and Render, $F(1, 60) = 0..07, p < .7920$; or the test score, Render, and Format, $F(2, 60) = 0.69, p < .5064$.

A Student-Newman-Keuls means test was performed *post hoc* on the main effect of Format. It indicated that at the alpha level of .05, the significant difference was between the Dynamic format and the two other formats, Sequentialand Static (see Figure 6). Subjects viewing the Dynamic format produced a significantly faster mean RT (M = 1.903 sec) than those viewing either the Sequential (M = 2.662) or Static (M = 2.332) formats. The difference between the RT for Static and Sequential was not found to be significant even though it was almost as large as the difference between Static and Dynamic.

The analyses indicated a significant effect on RT of the covariate, the Mental Rotations test score, F(1,60) = 9.46, $p < .0032$. The Mental Rotations test score accounted for 11.0% of the variability in RT. The removal of the test score as a covariate in the model showed a considerable drop in the predictive capability of the model ($R^2_{complete}$=.2823, $R^2_{reduced}$=.1703).

Error Rate

The ANOVA performed on error rate indicated a significant main effect for Render, F(1,60) = 13.53, $p < .0005$. The Render variable accounted for 15.1% of the variability in error rate. No significant interaction was found between the two independent variables of Format and Render, F(2, 60) = 0.10, $p < .9094$; the Mental Rotations test score and Format, F(2, 60) = 0.46, $p < .6323$; the test score and Render, F(1, 60) = 2.54, $p < .1163$; or the test score, Render, and Format, F(2, 60) = 0.48, $p < .6209$. The results of the ANOVA for error rate is summarized in Table 3. A comparison of the two levels of the Render variable revealed that subjects viewing objects with their edges represented as black lines had a significantly lower error rate (M = 1.472) than those viewing the shaded model (M = 2.917) (see Figure 7). Though there is a significant difference between the two levels of the Render variable, note that the mean score difference is only slightly more than a single incorrect response.

The analyses also indicated a significant effect of the covariate, the Mental Rotations test score, on error rate, F(1,60) = 8.29, $p < .0055$. The Mental Rotations test score accounted for 9.2% of the variability in error rate. The removal of the test score as a covariate in the model showed a drop in its predictive capability from $R^2_{complete}$=.2825 to $R^2_{reduced}$=.1882.

Table 2. Summary of RT ANOVA results

Source	**Df**	**Sum of Sq**	**Mean Square**	**F Value**	**Pr > F**
FORMAT	2	6.94870805	3.47435402	6.71	0.0024*
RENDER	1	0.66210219	0.66210219	1.28	0.2628
TEST	1	4.90231362	4.90231362	9.46	0.0032*
FORMAT*RENDER	2	0.26144858	0.13072429	0.25	0.7778
TEST*FORMAT	2	0.09099641	0.04549820	0.09	0.9160
TEST*RENDER	1	0.03634167	0.03634167	0.07	0.7920
TEST*FORMAT*RENDER	2	0.71306583	0.35653291	0.69	0.5064

* Significant at the .005 level

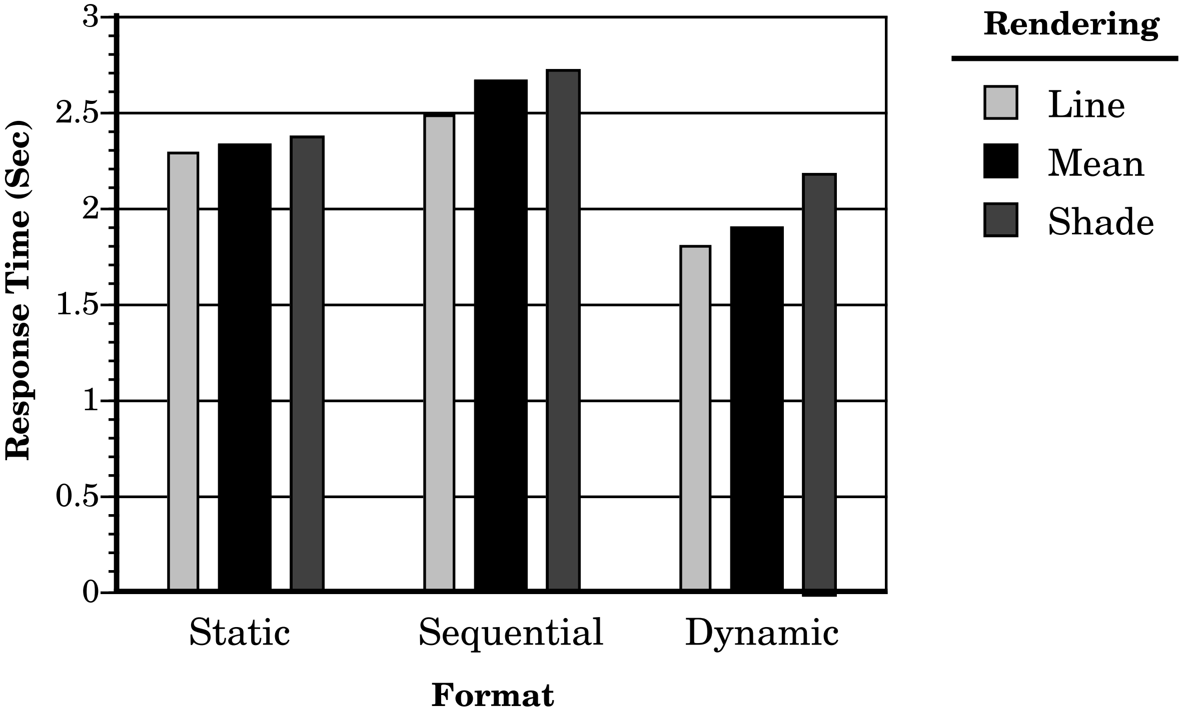


Figure 6. Summary of RT data

Table 3. Summary of Error Rate ANOVA results

Source	**Df**	**Sum of Sq**	**Mean Square**	**F Value**	**Pr > F**
FORMAT	2	9.36111111	4.68055556	1.69	0.1939
RENDER	1	37.55555556	37.55555556	13.53	0.0005†
TEST	1	23.01674198	23.01674198	8.29	0.0055*
FORMAT*RENDER	2	0.52777778	0.26388889	0.10	0.9094
TEST*FORMAT	2	2.56413489	1.28206745	0.46	0.6323
TEST*RENDER	1	7.04873660	7.04873660	2.54	0.1163
TEST*FORMAT*RENDER	2	2.66684252	1.33342126	0.48	0.6209

* Significant at the .01 level
† Significant at the .001 level

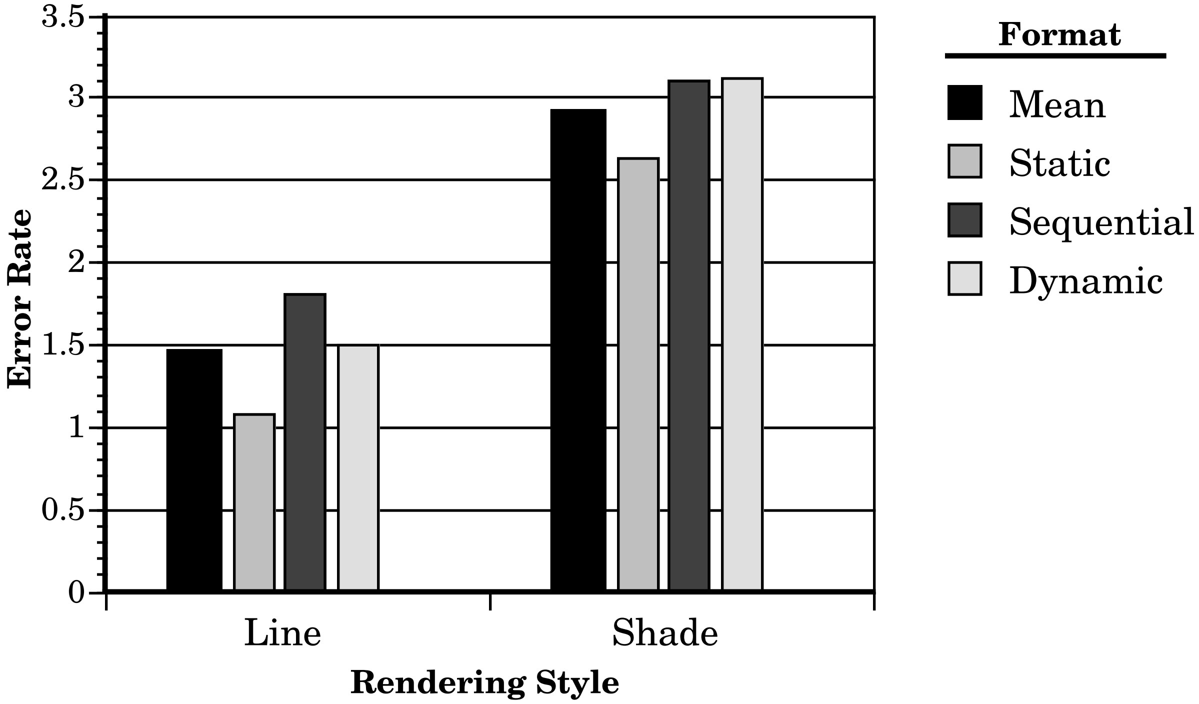


Figure 7. Summary of Error Rate Data

Discussion

As seen in the results of the ANOVA, both the Format and Render variables had a significant effect on different aspects of task performance. This result lends credence to the initial hypothesis that the design of a computer display representing a 3-D object is a factor in a subject's ability to identify that object from a forced choice pair. Inasmuch as the subject's task performance is correlated with the quality of his or her internal representation of the object, both the temporal dimension in which projections of an object were presented, and the method by which the edges and faces of the object were represented, affected the perception of the object. Though the results of the experiment don't confirm or disprove any particular theory of internal mental representation of 3-D form, they do indicate fruitful directions for future research into display designs.

The Format variable had a significant effect on the amount of time it took the subjects to choose one of the two objects from the forced choice pair. Those subjects seeing the Dynamic format responded significantly faster than those seeing either the Sequential or Static format. Though the difference was not significant, it was interesting to note that those subjects seeing the Static format responded faster than those seeing the Sequential format. This trend seems to rule out the inherent superiority of a serial presentation for this particular task. If a serial format were superior, then the Static format should have had the slowest Response Time.

With a more rapid presentation rate of projections, the Dynamic format differentiated itself from the Sequential format in two ways. First, the decreased time between projections allowed for the perception of apparent motion. Second, because each format was shown on the screen for the same length of time, the faster presentation of projections in the Dynamic format allowed for multiple presentations of the same projections within the same time period. Most likely both of these factors contributed to the reduced mean RT of subjects viewing the Dynamic format.

Unlike the Sequential format, the Dynamic format did not show the projections for a uniform length of time. Viewing an animated sequence induces a "momentum", that is, an expectancy of what the next projection will look like. Since the direction of rotation seen in the projection was reversing during the rocking back and forth motion, the animation "paused" at the three primary projections (right, front, and top) before reversing directions. This allowed the subject to perceptually prepare for the change in rotation direction. The rocking motion — first between the right and front and then between the front and top —, coupled with the pausing, created differences in both the number of times and total length of time a projection was seen (see Figure 8).

With the Static format, the subject had the capability of focusing attention on any projection, allowing a completely personalized strategy of viewing. With the complete Static display contained within a visual angle of approximately 29 degrees and individual projections within approximately 3.6 degrees, focus on either single projections or small groups of projections is possible. What was not possible with the Static projection, though, was the apparent motion present in the Dynamic format. With the Sequential format, no one projection was favored, as each was seen for only a single 1 sec presentation. This format had neither the flexibility of viewing of the Static format nor the apparent motion of the Dynamic format.

The number and length of time each projection was seen in the Dynamic format may be closer to an optimal distribution of favored views for forming the mental representation of the object. The right side, front, and top views are considered standard views in part because they tend to maximize the number of faces on a mechanical object (such as those used in this experiment) seen in their true size and shape. The shape of the objects used in this experiment coupled with the differences in formats described above meant that more time was spent viewing object faces in their true geometric proportions in the Dynamic format than in the Sequential format. The

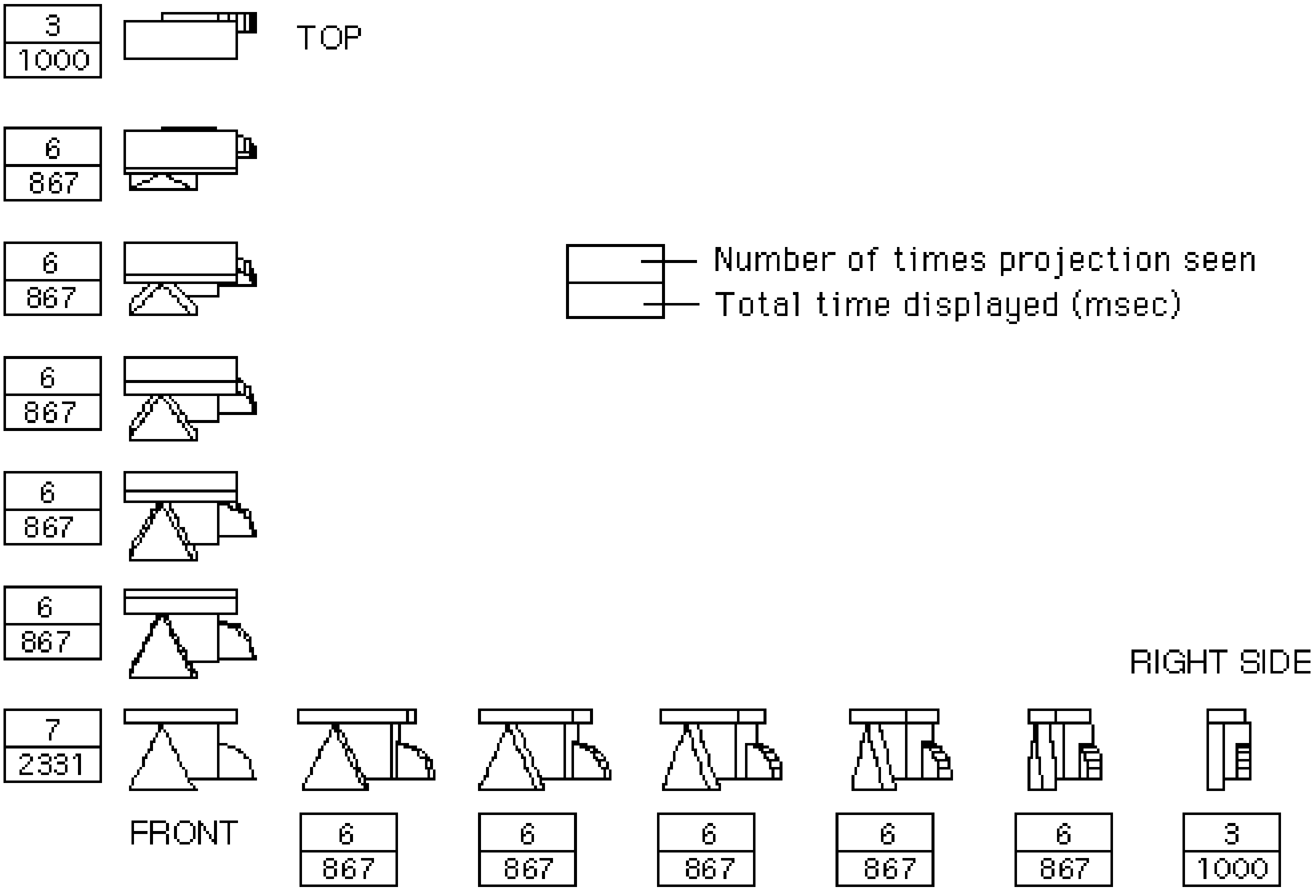


Figure 8. Length and frequency each projection was displayed in the Dynamic format

pausing at the end of each rock in one of the standard views allowed the subject to view a projection for an extended period of time and confirm connectivity of edges without having to track them through changing projections. This is one of the advantages the Static format affords. Unlike the Static view, the Dynamic also contained apparent motion which assisted in ordering of the edges and faces in depth. The combination of apparent motion and the ability to view the edges of faces in undistorted views combined to give the subjects viewing the Dynamic format the fastest mean RT.

These conclusions are supported in recent work by Sollenberger & Milgram (1993). Experimenting with a task of tracing paths in wireframe representations of blood vessels in the brain, the researchers concluded that simulating rotation of the structure on a computer display facilitated depth perception. At the same time, they concluded that continual rotation of the structure hindered the subject's ability the trace the connectivity of the lines representing vessels.

The Render variable had a significant effect on the number of times the subjects chose the wrong object of the pair, but not on RT. Subjects viewing the Shade rendering had a higher error rate than those viewing the Line rendering. These results differed from the results from a study by Barfield, et al. (1990) which found no significant difference between line and shade renderings in a task that involved judging the location of two objects in 3D space. One explanation may be the varying the representation of edges and faces with these different rendering methods may had more of an effect on perceiving spatial information *within* an object than *between* objects.

One possible explanation for the results seen in this experiment would be that the Shade renderings were more likely to lead to incorrect interpretations of the topology of the object. That is, the organization of edges and faces which combined to create the object. The problems in interpretation could arise either during the initial formation of the mental image of the object, in analyzing the forced choice pair, or in both subtasks.

Both the Line and Shade renderings of the objects represent edges where two surfaces of an object joined. The difference was in their method of representation. The Shade rendering attempted to more closely represent how these edges were perceived on real objects. Surfaces on a real object are perceived as having different values based in part on the orientation of the surface to the light source and the edge between surfaces is represented by a discontinuity in the gradient of value level. This line of discontinuity represents the juncture of two surfaces having different orientations to the light source and, therefore, to each other. The Line rendering abstracts this point of discontinuity by representing the edge as a line in a contrasting color to the surfaces.

In the case of this experiment, the Shade rendering represented surfaces in gradients of gray whereas the Line rendering depicted the edges as black and the surfaces as white. It could be that, though the Line rendering was a more abstract representation than the Shade rendering, it allowed a more direct recognition of edges and, therefore, was less prone to erroneous interpretation. The edges, in turn, were used in the perception of the whole object (Biederman, 1987). Another approach to looking at the differences in the performance of Line and Shade rendering is to apply the "no free lunch" axiom of the proximity compatibility principle (Wickens & Andre, 1990). The principle states that display information that is processed holistically enhances the speed of performance but decreases its accuracy. Display information which is processed in a focused, serial manner decreases the speed of performance but increases accuracy. A possible interpretation for the results of this experiment is that the Shaded format encouraged holistic processing, explaining the higher error rate. On the other hand, the RT data do not fully support this principle. Subjects using the Shade format consistently had slower RT than those using the Line format. The RT was, however, not significantly slower and further experimentation is probably needed to better test Wickens' and Andre's theory.

In contrast to this current experiment, Sanford, *et al.* (1987) found shade rendering superior to line rendering in a mental rotation task modeled after Shepard & Metzler (1971). Sanford, *et al.* did, however, note differences in performance depending on the shading algorithm used and on how many lighting sources were simulated in the rendering. These differences in rendering led to changes in the gradients of gray represented on the faces of the object. If the algorithms used for shading in the present experiment were substantially different than those used by Sanford, *et al.*, this may account for the differences in results. In addition, Sanford *et al.* worked with objects which were primarily tubular in shape. These shapes have substantially fewer discontinuities in the shading gradients representing edges, thus potentially fewer sources of misinterpretation of boundaries of surfaces.

The significant effect of the Mental Rotations test score as a covariate in the ANOVAs confirmed the initial assumption that the experimental task required mental imaging skills measured by this instrument. Mental Rotation Test score was significant in both ANOVAs, accounting for 11.2% of the variability of RT and 9.2% of the variability of the error rate. In contrast, neither the Render nor the Format variable was found to have a significant effect on both the dependent measures. Even where these variables had a significant effect on performance, they failed to account for the majority of the variability in performance. Format accounted for 15.5% of the variability of the RT and Render accounted for 15.1% of the variability of the error rate. Even when the variability explained by Render or Format is combined with the variability explained by the Mental Rotations Test score, a majority of the variability is still unaccounted for. The results of this study are a reaffirmation of the complexity of the perceptual and cognitive processes employed in the task simulated by this experiment. It is likely that the problem-solving strategies used by the subjects employ methods which were largely independent of the variables that were controlled in this experiment.

Conclusion

The results of this experiment show that manipulation of the temporal dimension of displaying projections of an object and how the object is rendered have a significant effect on the ability to identify the object in a forced choice pair. Along the temporal dimension, the dynamic display of projections of an object rotating in space led to the fastest RT. The Dynamic format provided the apparent motion of rotation yet allowed for confirming the connectivity of edges by pausing at key, standard views. Line rendering proved to give a lower error rate than Shaded rendering in the task. The indirect nature of representing edges with shading, the algorithms used to create the gradients on the faces, and the preponderance of planar faces on the objects used in this experiment all combined to increase the error rate. Future experiments could separate out these variables to more fully understand the role each one plays in the performance of this task.

The results of this study lead to the conclusion that tasks employing the projection of 3-D objects on 2-D computer displays need to be carefully matched to the method of displaying these projections. Both the complexity of the object and processing capability of the computer hardware may limit the speed at which projections of the object can be displayed. Depending on the task being performed, this may have a significant effect on performance. The use of line drawings rather than shaded renderings may allow for faster calculation and presentation of projections. The results of this study seem to indicate that this will not be detrimental to object identification tasks. The use of shaded objects needs to be carefully evaluated. Depending on the shape of the object and the shading algorithm employed, there may be markedly different performance levels. Because an object shaded in one software package on one system may look considerably different on other software or systems, the display needs to be evaluated in an operational setting.

The amount of variability accounted for by the two variables explored in this experiment is large enough to make both worth further study. One possible route would be to examine more carefully the sequence of projections viewed by the user. With proper equipment, viewers scanning the Static format could be analyzed for which projections were attended the most and in what sequence. Another related approach would be to use a serial presentation, but allow the user to interactively control which projection is viewed. The strategy employed by the viewer; including which projections were attended, the order in which they were viewed, and the length of that viewing, could be analyzed. Experiments using this design could confirm whether the standard views of right side, front, and top are favored when subjects have the freedom to choose projections. Manipulations in the shape of the object could be used in conjunction with these techniques to explore whether the subjects sought out views in which faces were seen in their true size and shape.

The use of interactive control of the projections could also be compared to the Dynamic presentation used in this experiment. A possible hypothesis is that active control of the displayed projections results in better performance in object identification tasks. This could be thought of as a comparison of passive and active viewing. Even restricting the format to a passive presentation, other formats could be explored. Having the object revolve a complete 360 degrees about each axis could be compared to the rocking motion shown in the Dynamic format. This would be another approach to isolating the importance of the three standard views. Rotating 360 degrees at a constant speed could give the perceptual cue of apparent motion without favoring any single view.

The Render variable could also be explored in more depth. There are many computer algorithms used for shading computer models. Many of them create different gradients of value across the surfaces of the object and these could be compared to the algorithm used in this experiment. Performance results in this experiment would be compared to the performance on displays in which the shader's clarity of the discontinuity represented edges between faces differed.

Another approach to evaluating the representation of edges would be to take a closer look at Line rendering. In addition to edges and faces, another perceived feature of 3-D form is the termination of edges at vertices. With the current Line rendering, edges are distinguished from surfaces by a strong value shift. Vertices could, in turn, be coded by altering the color of the ends of the edges where they terminate. This and other perceptual coding techniques recommended by Haber & Wilkenson (1982) could be used as a way of coding vertices. Various theories of perceptual organization (Biederman, 1987; Pomerantz, 1986; Treisman & Paterson, 1984) have discussed the role features such as edges and vertices play in the perception of three-dimensional form. The current experiment only looked at the broad categories of line vs. shaded rendering. A more systematic manipulation of the perceptual dominance of faces, edges, and vertices could be used to determine the importance of each of these elements in the identification of 3D objects. Of particular interest would be the interaction of various coding schemes in a dynamic rather than static display environment.

Bibliography

Allen, R. B. (1982). Cognitive factors in human interaction with computers. In A. Badre & B. Schneiderman (Eds.), Directions in human/computer interaction (pp. 1-26). Norwood, NJ: Ablex Publishing.

Baecker, R. M. (1987). Cognition and information processing theories. In R. M. Baecker & W. A. S. Buxton (Eds.), Readings in human-computer interaction: A multidisciplinary approach (pp. 299-307). Los Altos, CA: Morgan Kaufmann.

Barfield, W., Lim, R., & Rosenberg, C. (1990). Visual enhancements and geometric field of view as factors in the design of a three-dimensional perspective display. In Proceeding of the Human Factors Society - 34th Annual Meeting, (pp. 1470-1473). Santa Monica, CA: Human Factors Society.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115-147.

Braunstein, M. L. (1986). Perception of rotation in depth: the pyschophysical evidence. In N. I. Badler & J. K. Tsotsos (Eds.), Motion: representation and perception (pp. 242-247). Elsevier Science.

Eliot, J. (1987). Models of psychological space: Psychometric, developmental, and experimental approaches. NY: Springer-Verlag.

Gibson, J. J. (1966). The senses considered as a perceptual system. NY: Houghton Mifflin.

Green, B. F., Jr. (1961). Figure coherence in the kinetic depth effect. Journal of Experimental Psychology, 62, 272-282.

Haber, R. N., & Wilkenson, L. (1982). Perceptual components of computer displays. IEEE Computer Graphics &Applications, 2(3), 23-35.

Hochberg, J. E. (1964). Perception. Englewood Cliffs, NJ: Prentice-Hall.

Johansson, G. (1975). Visual motion perception. Scientific American, 232(6), 76-88.

Julesz, B. (1975). Experiments in the visual perception of texture. Scientific American, 232(4), 34-43.

Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propostions, and the form of internal representations. Cognitive Psychology, 9, 52-76.

Lappin, J. S. (1986). The representation and perception of geometric structure in moving visual patterns. In N. I. Badler & J. K. Tsotsos (Eds.), Motion: Representation and perception (pp. 79-92). Elsevier Science.

Lappin, J. S., & Fuqua, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. Science, 221, 480-482.

Lappin, J. S., Doner, J. F., & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. Science, 209, 717-719.

Marr, D. (1982). Vision. San Francisco: W. H. Freeman.

McCuistion, P. J. (1991). Static vs. dynamic visuals in computer-assisted instruction. Engineering Design Graphics Journal, 55(2), 25-33.

Paivio, A. (1983). The empirical case for dual coding. In J. C. Yuille (Eds.), Imagery, memory, and cognition (pp. 307-332). Hillsdale, NJ: Erlbaum.

Petersik, J. T. (1980). The effects of spatial and temporal factors on the perception of stroboscopic rotation simulations. Perception, 9, 271-283.

Pomerantz, J. R. (1986). Visual form perception: An overview. In C. Nusbaum & E. C. Schwab (Eds.), Pattern Recognition by Humans and Machines: Visual Perception Orlando, FL: Academic Press.

Sanford, J., Barfield, W., & Foley, J. (1987). Empirical studies of interactive computer graphics: Perceptual and cognitive issues. In Proceedings of the Human Factors Society - 31st Annual Meeting, (pp. 519-523). Santa Monica, CA: Human Factors Society.

Shepard, R. N. (1978). The mental image. American Psychologist, 33, 125-137.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701-703.

Sollenberger, R. L., & Milgram, P. (1993). Effects of Stereoscopic and Rotational Displays in a Three-Dimensional Path Tracing Task. Human Factors, 35, 483-499.

Teichner, W. H., & Mocharnuk, J. B. (1979). Visual search for complex targets. Human Factors, 21, 259-275.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. Journal of Experimental Psychology: Human Perception and Performance, 10, 12-31.

Tufte, E. R. (1990). Envisioning information. Cheshire, CT: Graphic Press.

Ullman, S. (1979). The interpretation of visual motion. Cambridge, MA: MIT Press.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills, 47, 599-604.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information display. Human Factors, 32, 61-78.

Wiley, S. E. (1990). Computer graphics and the development of visual perception in engineering graphics curricula. Engineering Design Graphics Journal, 54(2), 39-43.

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. Perception and Psychophysics, 31, 201-209.

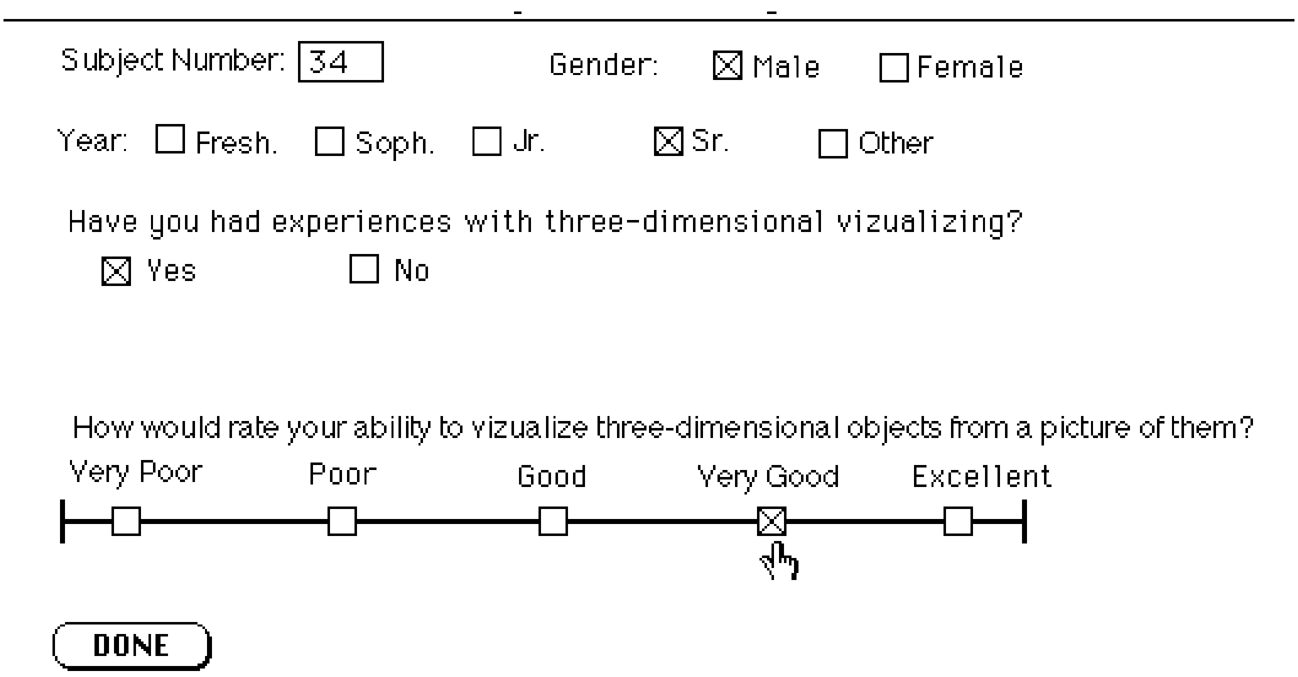
Zsombor-Murray, P. J. (1990). 2-D and 3-D CAD: complements to visualization. Engineering Design Graphics Journal, 54(3), 17-29.

Appendix A

Sample Questions From the Mental Rotations Test

Appendix B

Information Screen Seen Before Starting Stimuli



Subject Number: 34

Gender: ☒ Male ☐ Female

Year: ☐ Fresh. ☐ Soph. ☐ Jr. ☒ Sr. ☐ Other

Have you had experiences with three-dimensional vizualizing?

☒ Yes ☐ No

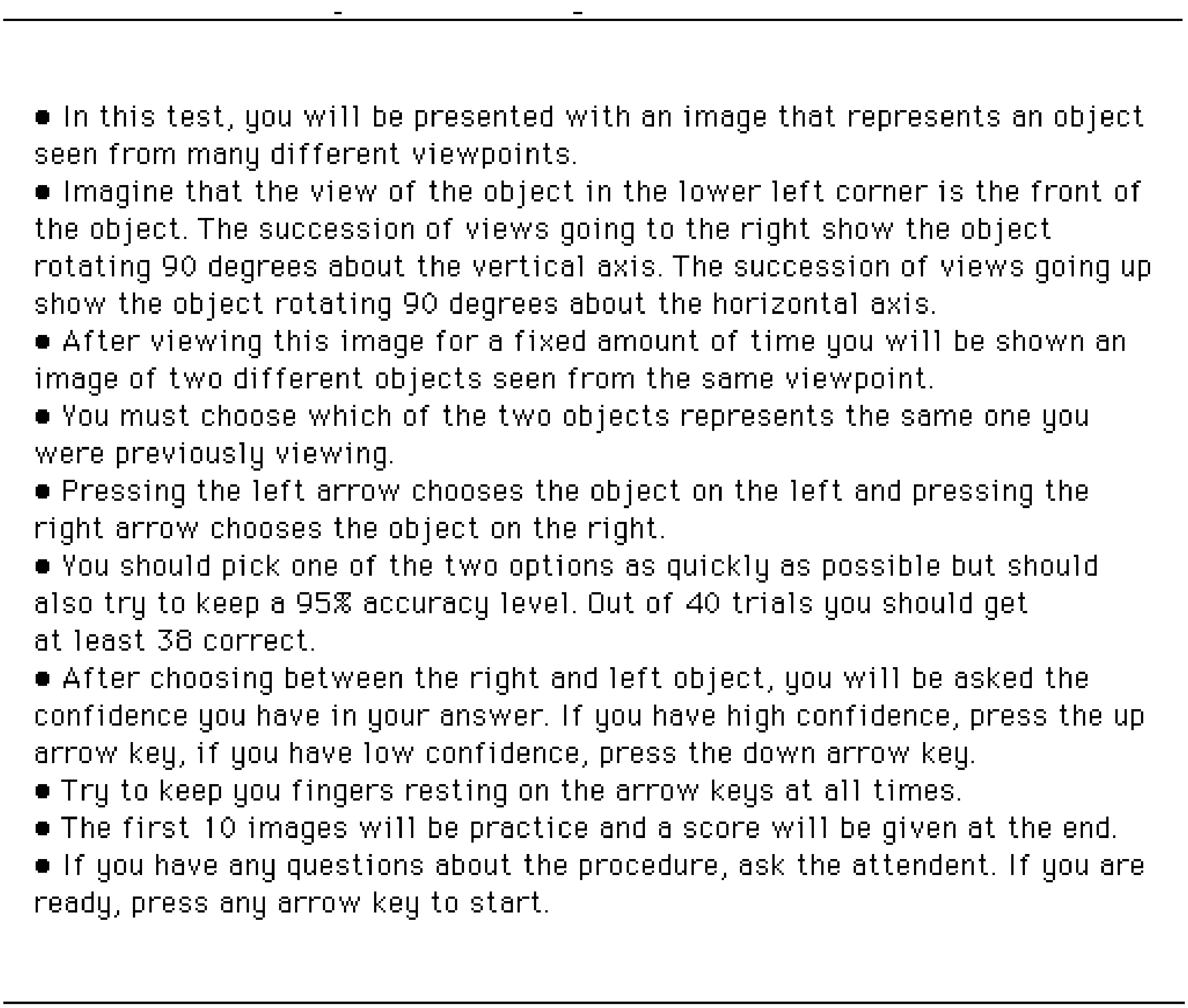
How would rate your ability to vizualize three-dimensional objects from a picture of them?

Very Poor ☐ Poor ☐ Good ☐ Very Good ☒ Excellent ☐

Appendix C

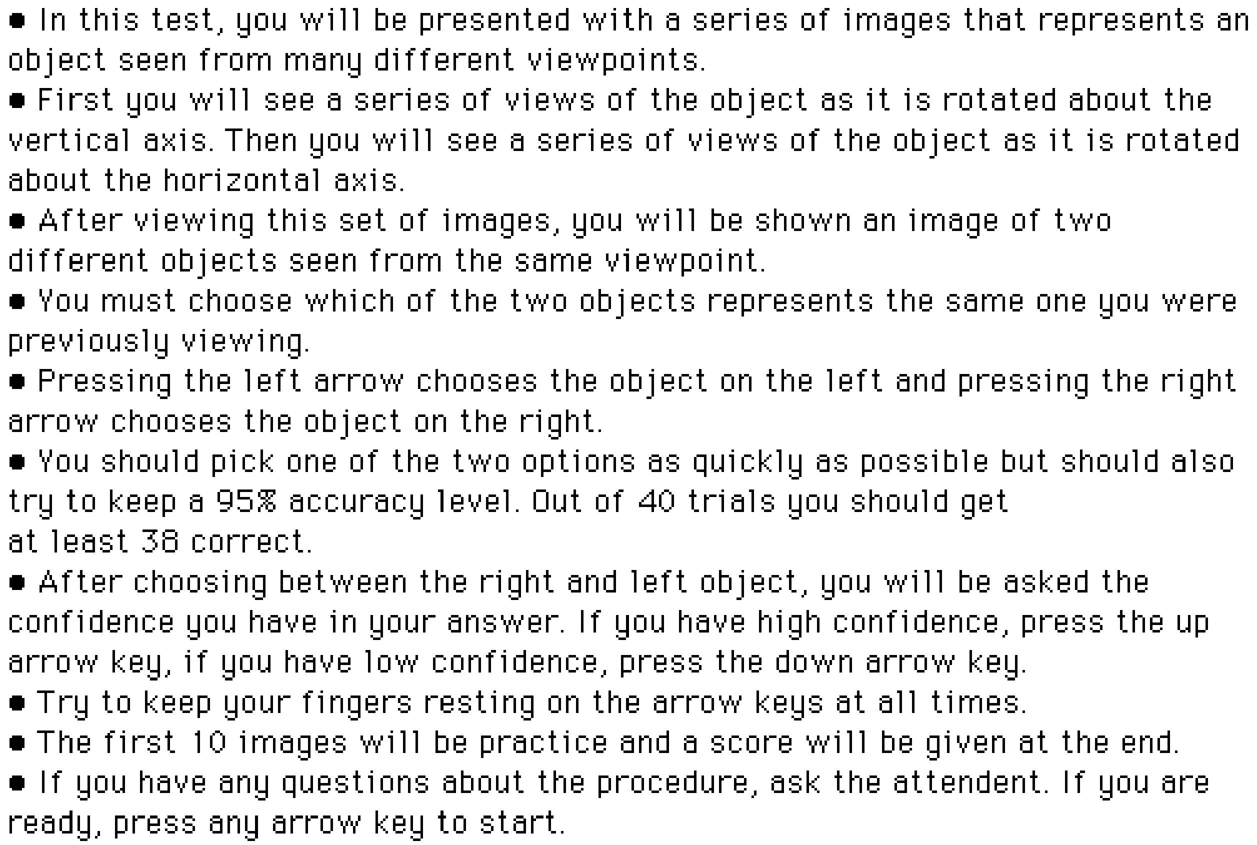
Instructions For Responding to Stimuli

Instructions for Static format



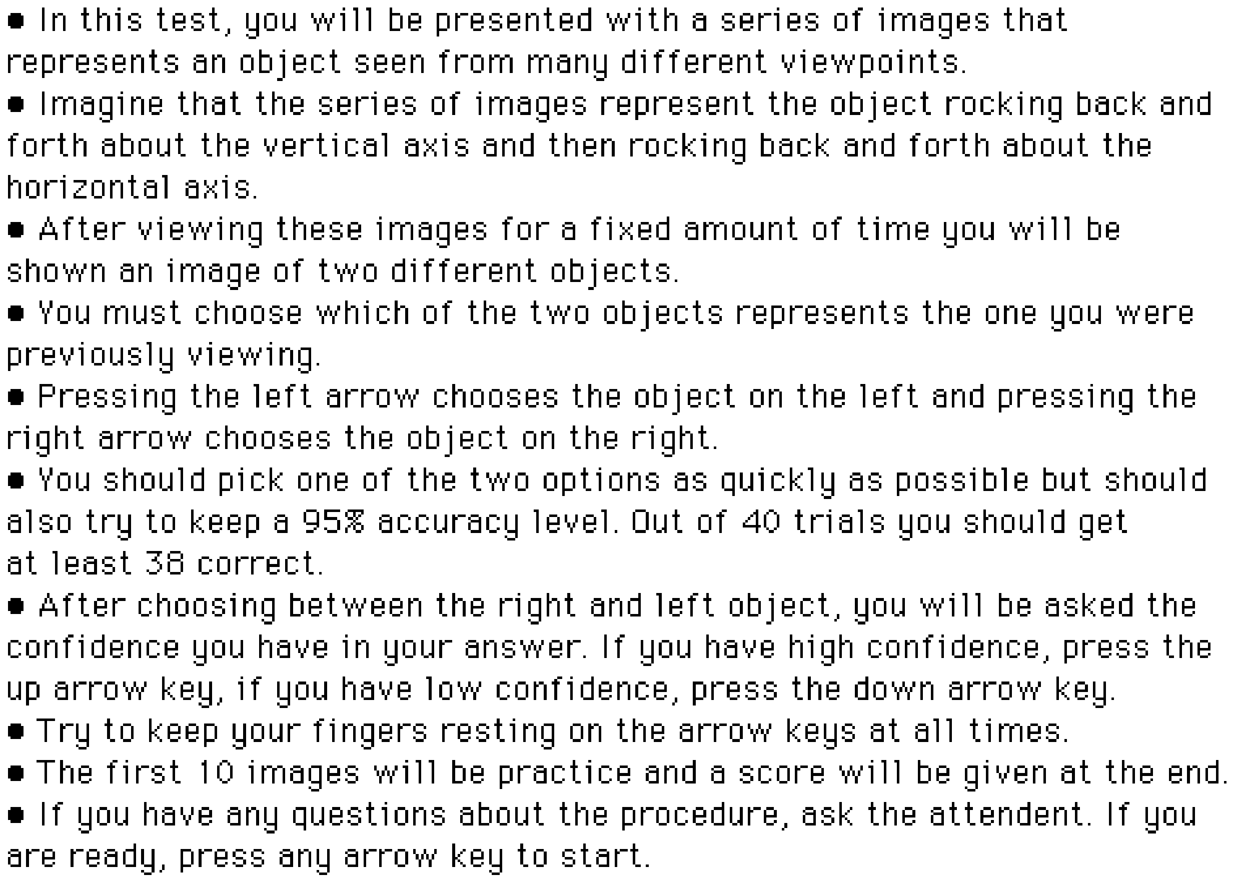
- In this test, you will be presented with an image that represents an object seen from many different viewpoints.
- Imagine that the view of the object in the lower left corner is the front of the object. The succession of views going to the right show the object rotating 90 degrees about the vertical axis. The succession of views going up show the object rotating 90 degrees about the horizontal axis.
- After viewing this image for a fixed amount of time you will be shown an image of two different objects seen from the same viewpoint.
- You must choose which of the two objects represents the same one you were previously viewing.
- Pressing the left arrow chooses the object on the left and pressing the right arrow chooses the object on the right.
- You should pick one of the two options as quickly as possible but should also try to keep a 95% accuracy level. Out of 40 trials you should get at least 38 correct.
- After choosing between the right and left object, you will be asked the confidence you have in your answer. If you have high confidence, press the up arrow key, if you have low confidence, press the down arrow key.
- Try to keep you fingers resting on the arrow keys at all times.
- The first 10 images will be practice and a score will be given at the end.
- If you have any questions about the procedure, ask the attendent. If you are ready, press any arrow key to start.

Instructions for Sequential format



- In this test, you will be presented with a series of images that represents an object seen from many different viewpoints.
- First you will see a series of views of the object as it is rotated about the vertical axis. Then you will see a series of views of the object as it is rotated about the horizontal axis.
- After viewing this set of images, you will be shown an image of two different objects seen from the same viewpoint.
- You must choose which of the two objects represents the same one you were previously viewing.
- Pressing the left arrow chooses the object on the left and pressing the right arrow chooses the object on the right.
- You should pick one of the two options as quickly as possible but should also try to keep a 95% accuracy level. Out of 40 trials you should get at least 38 correct.
- After choosing between the right and left object, you will be asked the confidence you have in your answer. If you have high confidence, press the up arrow key, if you have low confidence, press the down arrow key.
- Try to keep your fingers resting on the arrow keys at all times.
- The first 10 images will be practice and a score will be given at the end.
- If you have any questions about the procedure, ask the attendent. If you are ready, press any arrow key to start.

Instructions for Dynamic format



- In this test, you will be presented with a series of images that represents an object seen from many different viewpoints.
- Imagine that the series of images represent the object rocking back and forth about the vertical axis and then rocking back and forth about the horizontal axis.
- After viewing these images for a fixed amount of time you will be shown an image of two different objects.
- You must choose which of the two objects represents the one you were previously viewing.
- Pressing the left arrow chooses the object on the left and pressing the right arrow chooses the object on the right.
- You should pick one of the two options as quickly as possible but should also try to keep a 95% accuracy level. Out of 40 trials you should get at least 38 correct.
- After choosing between the right and left object, you will be asked the confidence you have in your answer. If you have high confidence, press the up arrow key, if you have low confidence, press the down arrow key.
- Try to keep your fingers resting on the arrow keys at all times.
- The first 10 images will be practice and a score will be given at the end.
- If you have any questions about the procedure, ask the attendent. If you are ready, press any arrow key to start.