Does Animation Affect Icon Learning and Recall?

Thesis

Submitted to

The College of Arts and Sciences of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree

Master of Arts in Psychology

by

Carlton D. Donahoo

UNIVERSITY OF DAYTON

Dayton, Ohio

December 2003

APPROVED BY:

CONCURRENCE:

ABSTRACT

DOES ANIMATION AFFECT ICON LEARNING AND RECALL?

Donahoo, Carlton, D.
University of Dayton, 2003

Advisor: Dr. David W. Biers

This laboratory investigation stems from an Air Force command and control research program in which the status of squadrons and individual aircraft in flight is communicated through pictorial icons. The study investigated the use of animation as a cue for remembering the status of aircraft presented in a 3 x 3 display of pictorial icons. Pictorial icons were presented to participants in both a homogeneous non-animated context (all non-animated) and a mixed context of animated and non-animated icons in a crossover design extending over two sessions. Results indicated that animation facilitated memory, but suggest that the effects may be short-lived. Animation is discussed in terms of its possible beneficial effects (saliency, isomorphism, highlighting distinguishing features, and information chunking) as well as its costs.

ACKNOWLEDGEMENTS

I am especially indebted to Dr. David Biers, my advisor, for the time, guidance, and patience that he provided throughout this thesis project. I would also like to express my appreciation to Dr. Greg Elvers and Dr. William Moroney for the helpful criticism that they provided for this project.

I am very grateful to everyone from the Human Factors Group at the University of Dayton Research Institute for the guidance and support that they provided. I would also like to thank everyone in the AFRL/HESR branch at Wright-Patterson Air Force Base for providing the equipment, facilities, and participants for this effort. Without their assistance I would not have been able to complete this project.

I would also like to extend my appreciation to my family and friends for the encouragement that they provided. I am especially appreciative to my grandmother, Fannie Ruffin, for the special encouragement, concern, and support that she supplied throughout this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

It has become commonplace for icons to be incorporated into computer interfaces. Icons are small pictorial symbols used to represent some sort of application, action, mode, or state. Most of the icons that are encountered within computer interfaces are static, i.e. they do not change their images over time. However, some static icons attempt to imply motion or action. These icons can be ambiguous or difficult to recognize because they try to convey motion in an unchanging format. In such situations it may be helpful to animate icons. By adding animation or motion to an icon, the underlying meaning of the icon may be better understood and its perceptual distinctiveness or saliency may be enhanced.

The Air Force Research Laboratory, Logistics Branch (AFRL/HESR) has been conducting research on a command and control program for wing level logistics group commanders called LOCIS (Logistics Control for Information Support). This program will allow Air Force wing commanders to view the status of their squadrons and individual aircraft in their command (Quill, Kancler, Revels, & Batchelor, 2001). The commander's display shows a visual account of the various aircraft squadrons on an airfield and in flight. Pictorial icons are used to represent the various aircraft within a squadron. These same icons also represent the condition or state of the aircraft. Some of these icons represent the aircraft when it is being prepared to become active (e.g., preparing for flight), while other icons represent the aircraft when it is active (e.g., in flight). The ability to determine the status of aircraft will assist wing commanders in making decisions regarding the aircraft.

As an aircraft gets closer to flight, the commander pays more attention to the aircraft's status. Icons, therefore, need to convey the state as unambiguously as possible. By adding animation to an icon, this ambiguity may be avoided by enhancing its perceptual distinctiveness and meaning. The notion of adding animation to some (or all) of the icons has been suggested by Air Force personnel as a means of improving the comprehension and discriminatability of these icons. The purpose of this study is to determine the impact of animation on the memory for the aircraft state of icons utilized in the LOCIS program.

Background

Axiomatic Design and the Need for Icon Animation in the LOCIS Program

In LOCIS, a commander must simultaneously consider a large amount of information along a number of different dimensions such as mission readiness, stage of aircraft readiness, position of aircraft in field, aircraft flight worthiness, and adherence to schedule. It is difficult to represent this information with simple uniform objects. However, by organizing the information along a number of visually salient dimensions, the user is presented with and may be better able to handle a greater amount of information than otherwise would be the case. Helander and Lin (2000) incorporated this principle into a design technique they titled "axiomatic design" in which distinct information requirements were mapped on to separate perceptually salient visual dimensions.

The foundation of axiomatic design is that functional information requirements (e.g., the need to show aircraft status) are matched to specific design parameters or "cues" (e.g. color) in order to "decouple" the design and avoid potential information conflicts (Helander and Lin, 2000). In other words, a single "cue" is mapped to a single functional information requirement. Table 1 presents the information attributes of the icons used in the LOCIS program and the cues used to represent them. When visual designs are coupled, users may have difficulty determining how to interpret certain information and to control specific functions. However, by pairing a given functional requirement with a specific visual cue, memory load may be reduced and users can thereby conduct a search more rapidly for a particular functionality. For example, the color of an aircraft may indicate whether or not an aircraft is flight worthy. If a user wanted to know how many aircraft were flight worthy they could do so by searching for the particular color associated with flight worthy aircraft. By searching for only that particular color they may conduct their search more quickly than if color was not used. This design technique was used in the development of the LOCIS display (Quill, et al., 2001).

In the LOCIS program, specific icons pictures have been chosen to represent the state of readiness of aircraft. However, given nine aircraft states and a small icon display area, some of icon pictures are not easy to discriminate. Since most of the typical perceptually salient visual cues have already been paired with functional requirements

Table 1

Coupling of Information Attributes and Visual Cues in the LOCIS Program

Cue	**LOCIS Information Attribute**
Color	Flight Readiness
Aircraft Picture	Stage of Readiness
View/Shape (Profile view vs. Top Down view)	On Schedule vs. Not on Schedule
Position	Position in real world
Shading (Filled in vs. Hollow)	Flyer vs. Spare

(e.g., color was paired with the flight worthiness of an aircraft), animation was chosen to be a cue to help determine which aircraft are ready to fly.

Icon Animation

Comprehension. Animation can be used to improve the comprehension of icons. The small pictorial symbols on icons are, generally, metaphors for some application, action, mode, or state. Lui (1997) stated that proper use of metaphors is important for creating effective graphical representations of abstract concepts and variables. Interfaces that use proper metaphors can make the strange become familiar, or make an abstract concept become concrete. Metaphors can help users relate a new unfamiliar situation to an old, familiar one so that they can use their existing knowledge for that new situation.

In general, the research on icons as metaphors has focused primarily on the use of static icons. For example, Wiedenbeck (1999) conducted a study in which static icons, text labels, and a combination of icons and text labels were compared in terms of learning a new computer task. The basic finding was that the combination of icons and text labels were the most effective for comprehension. However, there has been little research addressing the question of whether adding animation to an icon contributes to the user's ability to understand its meaning.

A major problem with iconic interfacing has been the ambiguity of the meaning of an icon. Lui (1997) states that a poorly selected metaphor or icon can be misleading if it supports alternative interpretations. Metaphors are also limited in their ability to explain ideas, so a perfect match may not exist for some operations. In other words, icons are more useful when the ambiguity between the actual icon and the icon's meaning is resolved.

Rogers (1989) addressed the problem of ambiguity in terms of iconic mapping. Direct mapping was defined as having the most similar structures between form and function. Icons that are the easiest to comprehend are considered to have the most direct mapping while the icons that are the most difficult to comprehend are believed to have the least direct mapping (Rogers, 1989).

According to Rogers, the effectiveness of an icon in relation to its intended meaning depends upon the degree of isomorphism between the physical form and function. For example a high degree of direct mapping would be represented by the use of a file folder to represent the data object of a file. There is an obvious isomorphic connection linking what is being represented and the form that is being used (Rogers, 1989). Less direct mappings represent their referents by drawing parallels between characteristics of the underlying referent and familiar structures in the icon. An example of an icon with a lesser degree of direct mapping would be the depiction of a pair of scissors to represent "to cut". This connection between form and referent takes the appearance of the most typical tool used to perform the act of cutting (Rogers, 1989).

Adding animation to an icon may make it more isomorphic to the underlying concept or process being depicted. For example, Baecker, Small, and Mander (1991) compared animated and non-animated icons in the redesign of a HyperCard tool palette. They used an iterative design process in which experienced users were exposed to increasingly refined iterations of the animated icons to remove ambiguity. Then novice users were presented the 18 tools within the palette in both animated and static forms. The results showed that the animated icons successfully convey the functionality of the HyperCard tools when the static versions of those icons failed to do so. Essentially, after the redesign, the animated icons were more obvious in their meaning to the novel users than the static icons. In another study, Bonder and MacKenzie (1997) compared the recognition accuracy for static and animated buttons representing concepts of varying degrees of complexity. Static and animated buttons were designed for 28 computer tasks. These tasks were grouped into low, medium, and high levels of complexity. Recognition accuracy was higher for the animated buttons than for the static buttons. It was also found that the animated buttons were more useful for representing high complexity tasks. This suggests that animation resulted in a more direct mapping for complex tasks than did its static representation.

Animated icons have the potential of more clearly and directly conveying their meaning to the system user than static icons when the referent itself involves motion. By adding animation to icons that represent some type of action or movement, their respective metaphors may be improved, thereby reducing or eliminating possible ambiguity (Alpert, 1991), In that the underlying concepts of some of the LOCIS icons represent movement or action, animation may be an important means of improving their comprehension through direct mapping to system referents. However, the effects of animation on learning and comprehension may be short lived. With sufficient training and sufficiently distinct icons, any relationship (even without animation) can be learned.

Icon Distinctiveness and Saliency. The commander's task in the LOCIS display involves scanning a large array of icons in varying states of aircraft readiness. Screen design techniques such as color usage, highlighting, and flashing have been used to attract users' attention to specific objects on a display (Tullis, 1997). However, little information has been gathered dealing with animation as a screen design technique for attracting visual attention. Techniques such as moving graphical objects on the screen or introducing time-based changes in the appearance of the screen elements may have an important impact on the user's interaction with the display (Tullis, 1997).

When there is an array of information from which to choose, the user must visually search these options to select the target. When computer screens are cluttered with all non-animated items, users usually need to spend a long time searching for the target. Distinct items can attract visual attention and can be found more quickly and more easily than non-distinct items. According to Wickens and Carswell (1997) an item is distinct if it differs from the other items on a least one physical attribute (e.g., amount of animation, size, brightness, hue, or rate of blinking). Animation should make the icons pop out if they are embedded in a field which also contains non-animated objects.

Treisman addressed the notion of making an item more salient for visual search. In her Feature Integration Theory, Treisman (1998) described the occurrence of "pop-out" in visual search. Essentially, items that have features that are different from features of surrounding items should be more distinctive or salient than its neighbors and appear to "pop out" of a display automatically. It could be argued that animation could be a feature to make an item more salient or stand out and grab the attention of a user. By making targets more salient than the items surrounding them, visual searches may be conducted more rapidly.

Although the LOCIS task is primarily one of visual search, it is possible that animation might also facilitate memory. Underwood (1969) proposed a theory of memory in which to-be-remembered items are tagged with attributes (e.g., imagery, affect, location). The attributes or tags can be used to make the memory more distinctive and facilitated recognition and recall. One such attribute that has been well-researched in both short-term and long term memory is that of imagery (Pavio, 1995). Perhaps animation would add imagery value to any to-be-remembered item in working memory and this imagery would facilitate memory by making the memory image more distinctive.

Amount of Animation, Differentiation, and Information Chunking. Animation may make icons "pop out" and be more salient. However, as suggested in the last section, the "pop out" effect is more likely to occur if the display contains a mixture of animated and non-animated icons. That is, if all icons are animated and moving, a particular icon may not stand out from the background. In addition, Alpert (1991) suggested that users may be distracted when all objects on a screen are moving. Alpert states that displays and interfaces that are composed of a mixture of animated and static icons may be preferable. Such displays may not be as distracting as a display full of animated objects, and may convey more information than a display completely composed of static objects. Therefore, based upon the above reasoning, the decision was made to automate only some of the LOCIS icons in the present study.

If this approach is to be used, there must be some logical assignment of icons to either static or animated representation. The icons chosen for animation were those that involved either underlying motion or those whose comprehension would be increase through animation.

Once the decision is made to animate only some of the icons, a new dimension is added to the task. Performance on the task could be enhanced merely by mentally grouping the icons into two sets- an animated and non-animated set-- and then searching for the icon within a particular target set. Thus, by chunking the information into two sets, the complexity of the overall task may be reduced through added icon class differentiation. If this is the case, then the presumed beneficial effect of animation may have nothing to do with increased saliency of the animated icons per se.

The Present Study

The present study was conducted in the context of the symbols being used to represent aircraft status in the LOCIS program. The study employed nine icons representing different aircraft states, (see Figure 1). Four of those icons were presented in both the animated and non-animated formats, and five icons were always presented in the non-animated format.

During the experimental phase, the participants viewed a 3 X 3 display of nine icons in various states of aircraft readiness. These icons could be in any one of the nine aircraft readiness states represented by the nine icon symbols. However, given only a 3 x 3 array to be searched, it was felt a visual search task would be too easy. In a real world setting such as LOCIS, the commander is presented a search task in which there are many icons (i.e. aircraft) which simultaneously vary along several attributes (e.g., icon symbol, color, shape, location). In the present study to gain experimental control, the number of icons was greatly reduced and varied along only two dimensions—icon readiness symbol and animation. Because of the simplicity of the task, it was felt that a visual search task would be too easy and not sensitive in picking up differences in search time as a function of animation. Therefore a memory task was used in which participants were shown a 3x 3 display for five seconds and then asked to correctly recall the aircraft

Icon	Aircraft State	Static Description	Animated Description	Blinking/ Movement
	Taxi	Aircraft with its landing gear down	Aircraft with its landing gear down and it is moving up and down.	Movement
	In-Flight	The aircraft is level just above the center of the icon.	The aircraft is flying in the sky with moving clouds in the background.	Movement
	Engine Start	The landing gear is down. Five lines representing exhaust are coming out of the engine	The landing gear is down. Five flashing lines representing exhaust are coming out of the engine	Blinking
	Pre-Flight	A fuel pump is attached to the aircraft. The canopy is open and landing gear is down	The fuel pump fills up (or hose blinks)	Blinking
	Crew Ready	The canopy is open. The landing gear and ladder are down.	No animation	No
	Crew Show	The canopy is open. The landing gear and ladder are down. One pilot is in the aircraft and the other is in the lower right hand corner.	No animation	No
	Take Off	The aircraft is tilted 45degrees toward the upper right hand corner.	No animation	No
	Landing	The aircraft is tilted 15 degrees towards the lower right hand corner and the landing gear is down.	No animation	No
	No Activity	Look at the aircraft from a bird's eye view.	No animation	No

Figure 1. Icons used in LOCIS program and the present study.

state of one of the nine aircraft after the display disappeared. In a visual search task all of the 9 icons would be visibly present and participant performance should be rapid. By using a recall task where the participants must use their memory, more of a demand would be placed upon the participant, and therefore result in greater sensitivity. However, the sensitivity was expected to be in percent correct rather than response time.

The experimental phase was conducted across two sessions, one in which all of the aircraft were presented in the non-animated format (homogeneous: non-animated context) and one in which four of the aircraft were presented in the animated format and five were always presented in the non-animated format (mixed context). The four icons that were animated in the mixed block were divided into two groups: those that involved flashing/blinking (Pre-Flight and Engine Start), and those that involved movement (Taxi and In Flight). The number of animated distracters in the mixed context was also varied. The primary measures were: (1) the percentage of correct responses, and (2) the response time.

The present study investigated five research questions related to animation. First, at a global level, does a mixed context of animated and non-animated icons result in greater recall of target aircraft state compared to a homogenous context of non-animated icons? Based upon the principle of differentiation, a mixed context may yield better recall than the homogeneous non-animated context because the mixed context the icons can be separated into two classes, animated and non-animated. By knowing the class to which the target belongs, the participant can reduce the memory load by limiting her/his memory search to the appropriate class. However when presented with the homogeneous non-animated, the participants must conduct a complete search of all of the icons because they are all similar in that all of the icons are non-animated.

Second, does animation of a target icon facilitate recall of the aircraft state over the non-animation of the same icon? This is the major research question addressed by the present study. If animation improves the comprehension of the underlying concepts through more direct mapping (Rogers, 1989), then recall of the animated targets should be facilitated. Alternatively, if animation makes the memory more distinctive as with mental imagery (i.e., stand out or pop out in memory), then one could expect the same beneficial effect of animation.

Third, does animation have any positive or negative effects on the recall of the non-animated targets? This is a key research question because it may shed some light on the alternative mechanisms (saliency vs. differentiation) for enhanced recall (if any) of the animated items. Because of the presumed saliency of the animated icons, it is possible that attention may be drawn away from the icons that are non-animated. If this is the case, recall of the non-animated icons would be less under the mixed than homogeneous non-animated condition because they were not processed sufficiently in the first place. Of course it is still possible that the mental images of the animated icons are more distinctive and therefore remembered better without having any adverse impact on the recall of the non-animated icons. On the other hand, animation may permit differentiation of the icons into two sets under the mixed condition. If this is the case, then recall of both the animated and non-animated icons should be enhanced under the mixed condition due to reduced memory load through chunking of information into two sets. That is, recall of the non-animated icons should also be greater under the mixed than homogeneous: non-animated conditions if differentiation is the sole mechanism by which animation operates.

Fourth, is there a difference in recall of icons as a function of the type of animation (flashing/blinking vs. moving)? There were two major types of animation used in the present study. One type of animation involved movement of the entire object (i.e., Taxi and In Flight) to make it isomorphic to the actual object that depicted motion. The second type of animation (flashing/blinking) was to make certain features of the icon more salient so that it easier to comprehend or easier to discriminate. Two of the icons in this study, pre-flight and engine start, had animation that was meant to increase the saliency of certain features of the icon and to make the features "pop out" to the user rather than to make the icon more isomorphic. There may be a difference in terms of the effect of animation based upon whether the animation was included to make the icon more isomorphic (movement) or certain features more salient (flashing/blinking).

Lastly, does the number of animated distracters affect the memory for the aircraft state of animated and non-animated targets? If animation attracts the participant's attention, then recall of the target icon (whether it be animated or non-animated) may decrease as the number of animated distracters increase.

CHAPTER 2

METHOD

Participants

The 20 participants (16 male, 4 female) were volunteer military (n=15) and civilian (n=5) logistics personnel from Wright Patterson Air Force Base, OH. All participants were very familiar with the flight line vernacular represented by the icons (e.g., take off, taxi).

Design

In the experimental phase of the study, the participants were presented a 3X3 display of nine aircraft icons, which could be in any one of the nine states represented by the nine icons. The participants' task was to correctly identify the aircraft state of one of the nine aircraft in the array. Participants were presented 162 test displays in each of two sessions under two different contexts--, one in which all aircraft were presented in non-animated states (homogeneous non-animated context) and one in which four of the nine icons were animated (mixed). Context was manipulated within-subjects with the order of presentation being counterbalanced. The homogeneous presentation of non-animated icons was chosen to represent the present state of affairs where all icons are non-animated. This provided a baseline for all comparisons. The set of mixed trials was selected to represent the near term plans of the military (LOCIS program) to animate some (but not all) of the icons.

There were 5 independent variables. The first independent variable was the Context in which the icons were presented (homogeneous non-animated vs. mixed). The second and primary independent variable was Icon Animation (animated vs. non-animated), which was manipulated within-subjects. The third independent variable, the Number of Animated Distracters (0, 1, or 2 animated distracters when the target was animated and 1, 2, or 3 animated distracters when the target was non-animated), was manipulated within-subjects and nested within the animation condition. The fourth independent variable was Type of Animation (blinking/flashing vs. movement), which was manipulated within-subjects only for the animated icons. The final independent variable was the Order in which the two experimental contexts were presented (homogenous non-animated context first and the mixed context second vs. mixed context first and the homogeneous non-animated context second). The dependent variables were the number of correct responses and response time.

Apparatus and Software

A 333 MHz IBM compatible, Pentium II computer with a 17 inch SVGA CRT monitor and QWERTY keyboard were used to interact with and display the testing matrices. The software to control the presentation of the stimuli in the training phase and the stimuli in both blocks of the experimental phase was written in Visual Basic 6. The software also collected the data from the participants' responses and stored that data in a Microsoft Access database.

Materials

There were nine icons with their associated labels. These nine icons were divided into two sets, an animated set and a non-animated set (see Figure 1 in Introduction). The animated set consisted of the Pre-Flight, Taxi, Engine Start, and In-Flight icons. The Pre-Flight icon was animated so that the fuel pump would fill up (or that the hose connecting the pump and the aircraft would blink). The Taxi icon was animated in that the aircraft moved up and down continuously. The animation for the Engine Start icon involved blinking/flashing lines representing exhaust flames. The animation for the In-Flight icon involved the aircraft flying through the sky with clouds moving in the background. These icons were chosen to be animated because they were difficult to recognize when not animated. Animated icons were grouped into two classes, those that involved blinking/flashing and those that involved movement. The Taxi and In-Flight icons involved movement, while the Engine Start and Pre-Flight icons involved blinking/flashing depending on which block in which they occurred. The animated set was presented in either the animated or non-animated state depending on whether the participants were under the homogeneous: non-animated or mixed context.

The non-animated set was always presented without animation. They consisted of the Crew Ready, Crew Show, Take Off, Landing, and No Activity icons. These icons were not animated because there would be too much distraction in the display if all icons were animated. The No Activity icon was considered a control icon and not analyzed in the data analysis to keep the number of icons equal in the animated and no-animated set. Each icon was displayed in a 64 by 64 pixel area.

Procedure

The experiment was divided into two phases, a training phase and an experimental phase. The experimental phase was further divided into two blocks of trials of 162 trials each. During one block (homogeneous non-animated context), all nine aircraft in a display appeared in the non-animated state. In the other block (mixed context) four of the nine icons were in the animated state when they appeared. The order of the two experimental blocks was counterbalanced across participants.

The study was conducted in two sessions. In Session 1, each participant was seated in front of a computer monitor and keyboard on their arrival for the experiment. Once seated the participant read along on the computer monitor as the instructions (see Appendix A) were read aloud to them. The instructions described the task that the participant was asked to complete. At the completion of this introduction the participant was asked to read and sign an informed consent form. Subsequently, participants engaged in training and in the first experimental block of trials under one of the two contexts. In Session 2, the experimental training and experiment phase were repeated, but under a different context. Once the participant completed the Preference Questionnaire at the end of Session 2 they were asked to read a debriefing form that described the purpose of the experiment.

Training Phase for Each Session

Before taking part in the experimental phase in each session, the participants were trained. The purpose of the training was to learn to recognize each aircraft symbol and its meaning. The participants received training under the same context as was used for the experimental block of trials for the given session. In other words, for the mixed experimental context, the participants trained with mixed set of icons and for the homogeneous non-animated experimental context the participants trained with all non-animated (homogeneous) icons.

A classic paired associate methodology was employed. A trial consisted of the presentation of each of the nine icon name pairings, one at a time. The participants were first shown each icon with its associated label one at a time for five seconds each. Then on each subsequent trial each icon was presented without its label. The participants' task was to correctly label the icon by pressing one of the function keys (F1-F9). Each key represented a different aircraft state label. The participant was required to respond within 2 seconds. If the participant responded correctly they proceeded to the next icon. However, if the participant responded incorrectly the word "Incorrect" appeared on the screen along with the icon with its correct label. If the participant did not respond within 2 seconds their response was considered to be incorrect and the correct pairing of the icon and its label was presented for an additional 2 seconds. If correct, the participant was then presented with the next icon. However if incorrect, the trial was terminated (due to a programming error). The training continued until the participant had correctly identified all of the icons on two successive trials. Upon completion of the training task, the participants began the experimental phase.

Experimental Phase for Each Session

The participants' task during the experimental phase of each session was to correctly identify the aircraft state of one of the nine aircraft. The aircraft were displayed to the participant in a 3X3 array (3 inches by 3 inches) with each aircraft being numbered 1 through 9. The participant initiated a trial by pressing the Enter key on the keyboard.

Upon pressing the Enter key, the 3X3 test array was displayed for five seconds. Then the test screen disappeared and was replaced by a blank screen. After 100 milliseconds the blank screen disappeared and was replaced with the response screen. In the response screen participants were shown a question asking them to identify the icon that had appeared in a specific position (i.e. Identify the icon in position X). Participants indicated their response by pressing one of the label function keys corresponding to the icon state. The participant had as much time as needed to give a response. After the participant responded feedback was given about the correctness of the response.

In each session, the participants were presented a block of 162 trials under one of the two contexts (homogeneous non-animated vs. mixed). In the 162 trials under each context, each icon was a target 18 times and occurred twice as the target in each of the nine aircraft positions.

The 18 trials for the given icon were first developed for the mixed context. A trial consisted of presentations of one target icon and eight distracters icons presented in the 3X3 test array. In each test array there was 1, 2, or 3 animated icons. Thus when the target was an animated icon, there was 0, 1, or 2 animated distracters. For the 18 trials of an animated target there were six trials each in which there were 0, 1, or 2 animated distracters with the restriction that no specific animated icon occur more than twice in any given test array. Any given animated target icon could occur twice in any given array (once as a target and once as a distracter). The non-animated distracters were chosen randomly with the restriction they occurred equally often over the 18 trials in the mixed context and occurring no more than twice in any array.

For the non-animated set of target objects in the mixed context, a similar procedure was used. For the 18 trials of a given non-animated target object, there were six trials each in which there were 1, 2, or 3 animated distracters. Non-animated distracters were chosen randomly with the restriction that they occur as equally often over 18 trials. The additional restriction was made so that no icon occurred more than twice in any given display (trial). The position of the distracter icons was determined randomly. Similar to the animated target set, any given icon could occur twice in any given array (once as a target and once as a distracter).

The above procedure generated 162 uniquely different test arrays or displays for the mixed context. The trial position of the 162 displays was determined randomly.

To generate the trials for the homogeneous non-animated context, a yoked procedure was used. The identical set of 162 arrays was used for the non-animated context with the exception that there was no animation. Thus there was an array in the mixed context that was identical to the array in the non-animated context, the only exception being the animation or non-animation of the icons.

Workload for Each Session

For each session, participants were administered the paper NASA TLX v1.0 workload assessment scale and asked to indicate their perceived workload on six subscales (see Appendix B). The participants' responses were converted to a 100 point scale in increments of five points per response category. After completion of the ratings, the participants engaged in a paired comparison task in which each pair of subscales was presented one at a time and were asked to indicate which member of the pair made a greater contribution to workload. The number of times each subscale was chosen was then used to formulate a weighted composite workload index. The presentation of results focus on the TLX composite.

Preference Questionnaire

At the conclusion of the second session, participants were asked to indicate their preference for the format (animated vs. non-animated) of the animated icon set (see Appendix C). For each icon in the animated set (taxi, engine start, pre flight, and in flight), participants were requested to place a check adjacent to the format they preferred (forced choice). The preference response measure was the number (percentage) of times the animated format was chosen.

CHAPTER 3

RESULTS

Presentation of the results is divided into two major sections—namely those for Session 1 and Session 2. This was necessitated by the significant interactions with Order for all analyses. Inspection of the data revealed that, in general, the effects were larger for Session 1 than Session 2. However, by treating each session separately, there is a corresponding loss of power to detect differences. This loss of power is attributable to the fact that the major independent variable (i.e., context) became a between-subject rather than within-subject factor and that the number of observations per condition was halved (i.e., reduced from 20 to 10).

Presentation of the results also focuses on correctness (i.e., trials to criterion, percent correct) rather than on latency in responding. Response latency was not a very sensitive measure in that there were very few significant differences with regard to response time when correct.

Session 1 (S1)

S1:Practice Trials - Learning of Icon-Name Pairings

The first task was to learn the name icon pairings using a paired-associate procedure. Remember that all icons were presented in the same state as would be experienced in the actual test phase of the experiment. This means that none of the icons were animated in the homogeneous condition whereas, in the mixed condition, the icons were split into two sets (animated and non-animated). Learning was continued until two successive perfect recitations were obtained.

The trials to criterion data revealed that when presented with the mixed context, participants took fewer trials ($M = 9.80$; $SD = 6.34$) to meet the criterion than when presented with the homogeneous non-animated context ($M = 13.70$, $SD = 8.69$). A one-way analysis of variance (ANOVA) was performed on the trials to criterion data. The results indicated that learning the icon-name pairings did not significantly differ as a function of context ($F(1, 18) = 1.31$, $p = .267$).

However, the aforementioned trials-to-criterion data may be misleading due to a procedural error. Rather than receive practice on every icon on every trial, a trial was terminated upon first error. This means that the icons were not practiced equally over the trials in which the subject participated. Therefore percent correct may be a better indicator of overall differences.

Table 2 presents the means and standard deviations for the percent correct measure for the overall data, and for the animated and non-animated subsets separately. As expected, overall and for the animated subset, the mean percent correct was greater under the mixed than homogeneous non-animated context. Note that there were no differences expected for the non-animated subset. However, the percentage of correct name-icon pairings did not significantly differ as a function of context overall ($F(1, 18) = 0.81$, $p = .380$), for the animated icon subset ($F(1, 18) = 0.55$, $p = .466$), or for the non-animated icon subset ($F(1, 18) = 0.98$, $p = .237$).

Table 2

Means and Standard Deviations for Percent Correct

During Learning Icon-Name Parings in Session 1

Set	Context: Homogeneous: non-animated		Context: Mixed	
	Mean	SD	Mean	SD
Overall	84.95	8.00	87.53	4.26
Animated	85.99	13.16	89.65	7.48
Non-Animated	85.73	6.62	88.73	6.97

S1:Experimental Trials - Testing the Memory for Target Icon State

S1:Research Question 1. The first research question addressed icon differences at a global level. Specifically the question addressed whether the mixed context of animated and non-animated icons would facilitate recall of the target icons compared to the homogeneous context of only non-animated icons. Results from the one-way ANOVA indicated that the effect of context ($F(1,18) = 11.03$, $p<.004$) was significant. The mean percent of recall for the target icons was significantly higher for the mixed context ($M = 61.67$, $SD = 9.83$) than the homogeneous non-animated context ($M = 48.27$; $SD = 8.12$).

S1:Research Question 2. The second research question focused on the animated icon subset—namely, would animation of a target icon facilitate recall of that icon compared to the non-animation of the same icon? The ANOVA supports the hypothesis, $F(1,18) = 19.72$, $p < .001$. There was significantly greater recall of the animated subset when animated ($M = 68.33$, $SD = 13.64$) than when non-animated ($M = 45.00$, $SD = 9.49$).

S1:Research Question 3. The third research question focused on the non-animated icon set. Specifically, would animation have any effect, either positive or negative, on the identification of a non-animated target icon? The results of the one-way ANOVA indicated that the effect of context ($F(1, 18) = 0.293$, $p = .595$) was non-significant. There was no evidence of animation having affected the memory of the target icon state for the non-animated icon subset with the means being 50.83 ($SD = 6.45$) and 55.33 ($SD = 13.10$) for the homogeneous non-animated and mixed context respectively.

S1:Research Question 4. The fourth question addressed whether there was a difference in the identification of an icon as a function of the type of animation (moving vs. blinking). To test this question a Type of Animation (moving vs. blinking) by Context (homogeneous non-animated vs. mixed) ANOVA was performed on the mean percent correct recall of the animated target set. Table 3 presents the means and standard deviations for these four conditions. There was no significant interaction ($F(1,18) = .768$, $p = .392$), indicating that there was no differential effect of the type of animation. However overall, there was better recall of the moving than blinking subset ($F(1,18) = 9.14$, $p = .007$). This indicated that recall of the two icons comprising the moving subset were easier to remember under all circumstances regardless of their animation state.

S1:Research Question 5. The fifth research question addressed whether the number of animated distracters (maximum of two) affected the identification of either animated or non-animated targets. In each case, a 3 (Number of Animated Distracters) by 2 (Context) ANOVA was performed on the mean percent correct recall of the target icons. It would be expected that there should be no effect of the number of animated targets in the homogeneous context since none of the icons were animated; however this was not the expectation in the mixed context. Therefore, if the number of animated distracters had an effect, there should be an interaction with context. The descriptive statistics for the animated targets are presented in Table 4. The results indicated that the interaction of context and the number of animated distracters for animated targets ($F(2, 36) = .214$, $p = .765$) was non-significant.

The effect of the number of animated distracters on the recall of non-animated targets is depicted in Table 5. Again, there was no-evidence that the number of animated

Table 3

Percent Correct Recall for Animated Icon Subset in Session 1

as a Function of Type of Animation

	Context					
Animation	Homogeneous: non-animated		Mixed		Overall	
Type	Mean	SD	Mean	SD	Mean	SD
Moving	50.56	13.66	71.39	12.97	60.97	16.80
Blinking	39.44	9.52	65.27	16.42	52.36	18.61

Table 4

Percent Correct Recall of Animated Targets in Session 1

as a Function of the Number of Animated Distracters

Number of Animated Distracters	Context: Homogeneous: non-animated Mean	SD	Mixed Mean	SD	Overall Mean	SD
0	45.55	12.30	70.83	21.48	58.21	21.40
1	44.00	10.96	69.75	14.73	56.88	18.28
2	42.03	12.40	63.99	11.82	53.01	16.31

Table 5

Percent Correct Recall of Non-Animated Targets in Session 1

as a Function of the Number of Animated Distracters

Number of Animated Distracters	Context: Homogeneous: non-animated Mean	SD	Mixed Mean	SD	Overall Mean	SD
1	47.17	10.80	57.50	14.67	52.33	13.61
2	48.75	11.34	49.53	11.37	49.53	11.37
3	55.83	3.44	55.68	12.65	55.68	12.64

distracters had any differential effect on recall--the interaction of context and the number of animated distracters ($F(2, 36) = 2.17$, $p = .134$) was also non-significant for non-animated targets.

Workload. At the end of the session, each participant completed the NASA TLX Workload Assessment instrument to measure the level of workload they experienced as a result of the experimental trials. As might be expected, the participants experienced less overall workload in the mixed context ($M = 77.25$, $SD = 12.29$) than in the homogeneous non-animated context ($M = 83.50$, $SD = 8.12$). However, a one-way between groups ANOVA indicated that the effect of context was non-significant ($F(1, 18) = 1.80$, $p = .196$).

Table 6 presents the subscale scores for the TLX under the two contexts. As shown in the Table, workload was higher under the homogeneous non-animated than mixed context for all but the Performance subscale. Somewhat surprisingly, only the Physical workload subscale revealed a significant difference as a function of context ($F(1, 18) = 5.51$, $p = .031$). Given that there were no apparent differences in the physical demands under the two contexts, these workload results are suspect.

Table 6

Subscale Scores for NASA TLX Workload Assessment in Session 1

Subscale	Context: Homogeneous: non-animated		Mixed	
	Mean	SD	Mean	SD
Mental	90.00	8.50	83.50	12.48
Physical	29.00	28.46	7.50	5.40
Temporal	87.50	9.79	76.00	25.03
Performance	61.00	30.26	71.50	15.82
Effort	83.00	16.70	72.50	21.25
Frustration	77.50	16.03	65.50	27.93

Session 2 (S2)

S2:Practice Trials - Relearning of Icon-Name Pairings in a Different Context

In Session 2, participants relearned the icon-name pairings, but in a different context (homogeneous: non-animated after mixed or mixed after homogeneous: non-animated). As might be expected relearning was much more rapid than original learning (Session 1) with the differences between the two contexts being smaller and in the same direction. The trials to criterion data revealed that when presented with the mixed context, participants took fewer trials (M = 3.20, SD = 1.48) to meet the criterion than when presented with the homogeneous: non-animated context (M = 4.40, SD = 2.80). However, the results indicated that relearning the icon name pairings did not significantly differ as a function of context ($F(1, 18) = 1.44$, $p = .246$).

Table 7 presents the means and standard deviations for the percent correct measure for Session 2 for the overall data and for the animated and non-animated subsets separately. As might be expected, percent correct for the icon-name pairs was very high (above 90%) for all conditions. Learning the icon-name pairings did not significantly differ as a function of context overall ($F(1, 18) = .448$, $p = .512$), or for the non-animated icon subset ($F(1, 18) = 1.36$, $p = .259$). It did significantly differ for the animated icon subset ($F(1, 18) = 6.37$, $p < .021$), with a higher percent correct in the homogeneous: non-animated context (M = 99.33) than the mixed context (M = 90.63). The means presented in Table 7 suggest a combination of two factors may have been operating to produce this difference for the animated subset of icons. Animation in Session 1 may have had a positive impact on the ease of relearning whereas experiencing something new during

Table 7

Means and Standard Deviations for Percent Correct

During Relearning Icon-Name Parings in Session 2

Set	Context: Homogeneous: non-animated		Mixed	
	Mean	SD	Mean	SD
Overall	95.07	4.38	93.16	7.90
Animated	99.33	2.11	90.63	10.70
Non-Animated	90.45	7.82	94.59	8.07

Session 2 (animation under the mixed condition) may have had a disrupting effect. The disrupting effect is supported by the larger standard deviation of the animated icon subset for the mixed condition.

S2:Experimental Trials - Testing the Memory for Target Icon State

Under a Different Context

S2:Research Question 1. Contrary to Session 1 data and opposite to the research hypothesis, the mean percent correct recall of the target icons was higher in the homogenous context (M = 61.17, SD = 17.09) than the mixed context (M = 58.03, SD = 15.75). However, results from the one-way ANOVA indicated that the effect of context ($F(1,18)$ = .184, p = .673) was non-significant.

When considering transfer from Session 1 to Session 2, there was little improvement in the percent correct recall for subjects going from the mixed (M = 61.07, SD = 9.83) to the homogeneous: non-animated (M = 61.17, SD = 17.09) context despite additional practice on the task. However, recall improved for those participants transferring from the homogeneous non-animated (M = 48.27, SD = 8.12) to the mixed context (M = 58.03, SD = 15.75). Without additional control groups (Homogeneous Non-Animated to Homogeneous Non-Animated and Mixed to Mixed), it is difficult to interpret these changes.

S2:Research Question 2. The second research question focused on whether animation of a target icon would facilitate recall of that target icon compared to the non-animation of the same icon. Although the mean percent recall of the animated subset was slightly higher under the mixed (M = 64.17, SD = 16.76) than homogeneous non-animated (M = 63.75, SD = 18.80) context, no significant effect for animation was found ($F(1, 18) = .003$, $p = .959$).

When examining changes from Session 1 to Session 2, there is some evidence to support the positive effect of animation within the animation subset. For the group of participants who transferred from the animated icons being in the animated (Mixed) to the Homogeneous non-animated state, percent correct decreased slightly from 68.33 to 63.75 percent. In the second group that transferred from the icons being in the homogeneous non-animated to the animated (Mixed) state, recall improved from 45.00 to 64.17 percent. During Session 1, there was clearly a benefit when the icons were animated. The fact that recall of the animated subset improved when animated during Session 2 is also consistent with the animation hypothesis. However, with lack of an appropriate control group, practice effects cannot be ruled out.

S2:Research Question 3. The third research question addressed whether animation would have a negative effect on the identification of a non-animated target icon. Consistent with the results of Session 1, a one-way ANOVA indicated that the effect of context ($F(1, 18) = .712$, $p = .410$) was non-significant. There was no evidence of animation having affected the memory of the target icon state for the non-animated icon subset with the means being (M = 55.56, SD = 18.79) for the homogeneous non-animated context and (M = 49.03, SD = 15.67) for the mixed context.

In comparing changes from Session 1 to Session 2, there was little variation in the recall of the non-animated subset unlike the animated subset of icons. For the group that transferred from the homogenous non-animated to the mixed condition, mean recall of the non-animated targets was 50.83 and 49.03 for the two sessions respectively. For the group that switched from the mixed to the homogeneous non-animated condition, the means were 55.33 and 55.56 respectively. Thus there was little evidence for improvement in task performance despite additional practice. This could be interpreted in terms of task difficulty or a possible detrimental effect of animation on non-animated targets.

Research Question 4. The fourth research question addressed whether there was a difference in the recall of an icon as a function of the type of animation (moving vs. blinking). As with Session 1, the moving set was easier to recall overall ($F(1, 18) = 5.34$, $p = .033$). However, unlike Session 1, the interaction for the type of animation and context ($F(1, 18) = 18.41$, $p < .001$) was significant. As shown in Table 8, the significant interaction represents a reversal in the direction of the effect. Namely, animation facilitated recall of the moving subset, but inhibited recall of the blinking subset. Despite the overall significant interaction, analysis of simple effects revealed that the effects of animation was not significant for either the moving ($F(1,18) = 1.52$, $p = .233$) or the blinking ($F(1,18) = .02$, $p = .901$) subset. Thus, taken together with the results of Session 1, there was little support that the type of animation differentially affected the outcome of the study.

Table 8

Percent Correct Recall for Animated Icon Subset in Session 2

as a Function of Type of Animation

	Context					
Animation	Homogeneous: non-animated		Mixed		Overall	
Type	Mean	SD	Mean	SD	Mean	SD
Moving	61.67	19.19	71.11	14.77	66.39	17.36
Blinking	65.85	18.80	57.22	20.42	61.53	19.61

S2:Research Question 5. The fifth research question addressed whether the number of animated distracters affected the recall of either animated or non-animated targets. Tables 9 and 10 present the mean recall as a function of the number of distracters for the animated and non-animated targets respectively. Consistent with Session 1, the results indicated that the interaction of context and the number of animated distracters was not-significant for either animated ($F(2, 36) = .333$, $p = .719$) or non-animated ($F(2, 36) = .007$, $p = .993$) targets.

S2:Workload. Participants perceived the workload to be 5-10 % less than in Session 1 with the workload differences being smaller but in the same direction. The mean workload on the NASA TLX was 72.20 ($SD = 14.98$) for the mixed context and 73.53 ($SD = 13.79$) for the homogeneous: non-animated context. Table 11 presents the subscale scores for the TLX under the two contexts. As with Session 1, the results indicated that the effect of context ($F(1, 18) = .043$, $p = .838$) was non-significant. Table 11 presents the subscale scores for the TLX under the two contexts. As shown in the Table, workload was higher under the mixed than homogeneous non-animated context for all but the Performance and Frustration subscale.

Preference

At the conclusion of the study, all participants were asked to indicate their preference for the icon format (animated or non-animated) of the four icons of the animated set in a forced-choice situation. Results were tabulated in terms of the percentage of choices for the animated and non-animated formats. Using a one-sample t test with chance being set at 50%, there was a significant preference ($t(19) = 2.90$, $p = .009$) for the icons when animated ($M = 65.50$ %) than when non-animated ($M = 34.5$%).

The preference for the animated version of the icon was significant for the moving type of animation [t (19) = 2.93, p = .009] (M = 72.50%) but not for the blinking type of animation [t (19) = 1.56, p = .135] (M = 62.50%).

Table 9

Percent Correct Recall of Animated Targets in Session 2

as a Function of the Number of Animated Distracters

Number of Animated Distracters	Context: Homogeneous: non-animated Mean	SD	Mixed Mean	SD	Overall Mean	SD
0	66.83	20.53	66.83	16.48	66.83	18.12
1	64.42	20.22	64.58	16.88	64.50	16.88
2	58.51	20.61	63.39	19.39	60.95	19.63

Table 10

Percent Correct Recall of Non-Animated Targets in Session 2

as a Function of the Number of Animated Distracters

Number of Animated Distracters	Context: Homogeneous: non-animated Mean	SD	Mixed Mean	SD	Overall Mean	SD
0	54.58	20.27	48.83	23.09	51.46	21.39
1	54.29	19.22	48.57	16.37	51.43	17.62
2	58.17	22.04	52.62	14.74	55.39	18.47

Table 11

Subscale Scores for NASA TLX Workload Assessment Session 2

	Context			
Subscale	Homogeneous: non-animated		Mixed	
	Mean	SD	Mean	SD
Mental	80.50	12.57	84.50	12.12
Physical	10.50	11.41	27.50	27.31
Temporal	71.00	24.47	80.50	13.83
Performance	64.50	16.57	54.50	20.61
Effort	76.00	14.87	83.00	12.73
Frustration	54.50	31.84	47.00	30.48

CHAPTER 4

DISCUSSION

The present study investigated the effects of animation on the memory for the aircraft state of icons used in the LOCIS program. This study provides some evidence that animation of pictorial icons can lead to improved recall of target objects. Support for animation as a retrieval cue was evidenced by the better overall recall under the mixed than homogeneous non-animated context (Research Question 1) and by the better recall of the animated set of icons when animated than when non-animated (Research Question 2). However the above significant effects of animation were limited in that they were present only in Session 1. Despite failure to replicate these effects of animation on memory in Session 2, participants showed a preference for the animated version of the animated set over the non-animated version, and this was particularly true for the icons comprising the moving type of animation.

It should be noted that when the Session 2 results are viewed in perspective to Session 1, there was some additional support for the positive effect of animation within the animated subset. The participants that transferred from the homogeneous non-animated context to the mixed context experienced an improvement in recall of the animated subset. This was not true for the non-animated subset of icons.

No statistical evidence was found to indicate that animation affected the recall of the non-animated subset of icons (Research Question 3). The type of animation (blinking or moving) did not statistically affect memory for target icon state (Research Question 4). The number of animated distracters did not reliably affect recall for either animated or non-animated targets (Research Question 5). Finally, animation of the animated icon set did not significantly affect the learning and comprehension of the icons.

A Caveat: Order Effects and Consequences for Power

When designing the present study, it was felt that the crossover design (counterbalancing the order of the within- subject manipulation of context) and a sample size of 20 permitted sufficient power to detect most treatment effects. It was unanticipated that there would be interactions with order. However, with significant interactions with order and the necessity to analyze each session separately, there was a loss in sensitivity in the present study—i.e., the sample size was halved, and the context manipulation became a between-subjects variable. The reason for mentioning the above is that several of the differences were in the presumed direction, but just not significant. For example, workload was lower under the mixed than homogeneous: non-animated context. Other effects are discussed in the following sections.

Possible Explanations for Animation Effects

Comprehension and Learning

The icons chosen for animation were those that involved either underlying motion (isomorphic to actual state) or those whose comprehension would be increased through animation. Therefore it was anticipated that association between the icon and its meaning (i.e., aircraft state) would be learned more rapidly in their animated than non animated state. There is some support for this possibility, albeit minimal. The trials-to-criterion data in the paired-associated training phase revealed that the icon-name pairings were learned more rapidly under the mixed than homogeneous non-animated context, although the difference was non-significant. More importantly, although again non-significant, the percent correct measure indicated that the learning the icon-name pairings for the animated set was higher when the icons were animated (mixed context) than when non-animated (homogeneous: non-animated context). These differences, however, may have been underestimated. Remember that due to a procedural error, a trial was terminated upon first error. This omitted the chance to verify the correctness of items already learned and to practice the correct alternative.

Despite the above, the percentage of correct icon-name pairings for the non-animated icon set was also higher under the mixed than homogeneous non-animated context (non-significant but by the same magnitude). This suggests that improved learning under the mixed context, if it did occur, may have been due to some sort of differentiation of the icons rather than increased comprehension of the animated set per se.

Icon Differentiation, and Information Chunking

When animation occurred in the present study, it was always in the mixed context. Perhaps this context allowed participants to better focus their attention. They could better remember the target because they could chunk the information into two distinguishable subsets of icons—those animated and those not animated—and this chunking reduced the memory load. Perhaps the same phenomenon would not have occurred had all icons been animated.

Although memory for animated subset of icons in the experimental phase was facilitated in the mixed context (Research Question 2), the memory for the non-animated subset was not affected (Research Question 3). This suggests that the better recall in the mixed context (Research Question 1) is not simply a matter of reduced memory load through chunking of information. If it was, the expectation would be that the recall of the non-animated set would have been facilitated under the mixed context as well.

Icon Distinctiveness and Saliency

Animation may make icons "pop out" and be more salient. However, the "pop out" effect is more likely to occur if the display contains a mixture of animated and non-animated icons (i.e. mixed context). The present data are consistent with this hypothesis in that recall of the animated subset was facilitated when animated (mixed context) (Research Question 2) whereas recall of the non-animated subset was not affected by context (Research Question 3).

Triesman (1998) in her feature integration theory of visual search argues that items that have features different from the surrounding items should be more distinctive or salient than its neighbors and appear to "pop out" of a display automatically. It could be argued that animation could make an icon more salient or grab the attention of the user. Perhaps the same phenomenon operates in memory. Underwood (1969) proposed a theory of memory in which to-be-remembered items are tagged with attributes (e.g., imagery, affect, location). These attributes or tags can be used to make the memory more distinctive and facilitate recognition and recall. One such attribute that has been well researched in both short- and long-term memory is that of imagery (e.g., Paivio, 1995).

Perhaps animation adds imagery value to a to-be-remember icon in working memory and this imagery facilitates memory by making the memory image more distinctive.

Effects of Animation on Non-Animated Icons (Research Question 3)

Concern was manifested in the Introduction over the possible effects of animation on the memory for non-animated objects. Because of the presumed saliency of the animated icons, it was conjectured that possibly animation would draw attention away from the non-animated icons and interfere with memory. However, this was not the case in that recall of the non-animated icons did not significantly differ in the two contexts.

The failure to find a significant effect of context for the non-animated icons also argues against differentiation as the sole explanation for the context effect. If differentiation was the explanation for the context effect then recall of the non-animated set of icons should have been better under the mixed than homogeneous: non-animated context. It was not better.

Type of Animation (Research Question 4)

The present study also investigated if the type of animation (flashing/blinking vs. moving) has any influence on recall of the target icon. In two of the cases, animation (Taxi and In Flight) involved adding movement to the depicted object to make it isomorphic to the actual motion of the underlying concept (Rogers, 1989). In two other cases (Pre-Flight and Engine Start), animation involved the addition of blinking to make critical distinguishing features stand out. Although there was no evidence to support any differential memory effects of the two types of animation, the moving set was remembered better than the blinking subset in both their animated and non-animated state. This main effect simply indicates that the two icons belonging to the moving subset (Taxi and In Flight) were easier to remember regardless of their animation state.

Interestingly, although there were no differential memory effects of the two types of animation, participants preferred the animated version of the moving subset over the static version whereas no such evidence was found for the blinking subset. Preference for the animated subset is consistent with the principle of isomorphism.

The Number of Animated Distracters (Research Question 5)

Alpert (1991) suggested that users may be distracted when all objects on a screen are moving. Alpert stated that displays and interfaces that are composed of a mixture of animated and static icons may be preferable. Based upon an extension of this reasoning, it was conjectured that as the number of animated distracters increased, recall of both animated and non-animated targets would decrease. There was no support for this hypothesis in either the pattern of the means or significance. One possible reason for failure to find differences as a function of the number of distracters is that the number of distractions varied in a narrow range (0 to 2 for the animated set and 1 to 3 for the non-animated set).

Considerations for Future Research

The data from Session 2 suggests that the animation effects may be short lived. However given the crossover design and the significant interactions with order, the results of the Session 2 data cannot be unambiguously interpreted without additional control groups (i.e., homogeneous non-animated to homogeneous non-animated, and mixed to mixed). The differences in Session 2 results could have been due to the specific

Session 2 context (homogeneous non-animated vs. mixed), prolonged practice, or the specific context of initial learning affecting transfer to the other context. Only the results from Session 1 are clearly interpretable. Therefore, the study needs replication with the addition of the above two control groups (homogeneous non-animated to homogeneous non-animated, and mixed to mixed).

The present study used a memory task rather than a search task. A search task would be more representative of the actual commander's task when viewing the LOCIS display. The decision to use a memory task in the present study was based upon the feeling that the simplicity of the 3 x 3 display would not be sensitive to potential differences in search time. The study needs replication using a search paradigm in conjunction with a more complex search matrix.

Implications for the Design Community

Finally, although the present study provides some evidence for the beneficial effects of animation, the design community has to ask the question: at what cost? Animation is more expensive in terms of the involving more time to implement (e.g., program), and may affect the update speed of the display. There are other perceptually salient cue dimensions (e.g., color, shape, position) that provide greater benefit at less cost. However, in the current application where more traditional cues have been exhausted, animation provides an acceptable alternative. The major remaining question for the LOCIS command and control application environment is whether the present results would hold in a visual search recognition paradigm.

REFERENCES

Alpert, S.R. (1991). Self-describing animated icons for human-computer interaction: A research note. *Behaviour and Information Technology, 10*, 149-152.

Baecker, R., Small, I., & Mander, R. (1991). Bringing icons to life. *Proceedings of the CHI 91' Conference on Human Factors in Computing Systems,* 1-6.

Bonder, R.C., & MacKenzie, I.S. (1997). Using animated icons to present complex tasks. *Proceedings of CASCON'97,* 281-291.

Helander, M. G., & Lin, L. (2000). Anthropometric design of workstations. *Proceedings of ICAD2000 First International Conference on Axiomatic Design,* 1-9.

Lui, Y. (1997). Software-user interface design. In Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1689-1719). New York, NY: John Wiley & Sons Inc.

Paivio, A. (1995). Imagery and memory (pp. 977-986). In Michael S. Gazzaniga (Ed.), *The Cognitive Neurosciences*; Boston, MA: MIT Press.

Quill, L. L., Kancler, D. E., Revels, A. R., & Batchelor, C. (2001). Application of information visualization principles at various stages of system development. *Proceedings of the Human Factors and Ergonomic Society 45th Annual Meeting,* 1713-1717.

Rogers, Y. (1989). Icons at the interface: Their usefulness. *Interacting with Computers, 1*, 105-117.

Treisman, A. (1998). The perception of features and objects. In Wright, R. D. (Ed.), *Visual Attention* (pp. 26-54). New York, NY: Oxford University Press.

Tullis, T. S. (1997). Screen design. In Helander, M. G., Landauer, T. K., & Prabhu, P. V. (Eds.), *Handbook of Human-Computer Interaction* (pp. 503-531). Amsterdam, The Netherlands: Elsevier.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*, 559-570.

Wickens, C. D., & Carswell, C. M. (1997). Information processing. In Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 89-129). New York, NY: John Wiley & Sons Inc.

Wiedenbeck, S. (1999). The use of icons and labels in an end user application program: An empirical study of learning and retention. *Behaviour & Information Technology, 18*, 2, 68-82.

APPENDIX A

Instructions, Consent, and Debriefing Forms

Instructions

Welcome to the "Icon Animation" study. My name is Carlton and I will be administering the experiment. Please follow along on the instruction sheet in front of you as I read the instructions aloud. Also, please feel free to ask any questions that you have as we go through the instructions.

Before we begin I need you to read over and sign an informed consent form. This includes a brief description of the study and your rights as a potential research participant. Please let me know if you have any questions after reading this form, otherwise please sign and date both copies.

This study will require you to complete two days of testing. On the first day of testing you will complete session one and on you second day of testing you will complete session 2. For this study you will be asked to complete three or four general tasks (depending on whether it is your 1st or 2nd day of testing): view a series of icons and identify specific icons during a training task, view a series of 3x3 displays that are composed of the icons encountered in the training phase and identify specific icons, complete the NASA TLX workload assessment scale, and answer a post-test questionnaire after completion of both sessions. A researcher will be available throughout the experiment; however, the researcher will not be able to assist you during the experimental tests.

For the first task you will be asked to complete the necessary training for the study. When you are ready to begin this task you can click on the "Start" button located in the middle of the screen. You will then see an aircraft icon and its name. After two seconds the icon will disappear. At this point you should click on the "Next" button. A "Previous" button will appear to the left of the "Next" button. This "Previous" button can be used to view the icon/name pairings that you have reviewed. After you have reviewed all of the icon/name pairings you will be given the option to either proceed to the Pre-Test or repeat the review of the pairings. To make this selection either click on "Previous" button or "Done" button. By clicking on the "Previous" button you will move backwards to the last icon/name pairing you had viewed. Clicking on the "Done" button will bring you to a screen that will again ask you if you would like to repeat the icon/name review. By clicking on the "Yes" button you will repeat the review. By clicking on the "No" button you will proceed to the Pre-Test. At this time you may begin the review session. Please notify the experimenter once this task has been completed.

During the Pre-Test you will be asked to correctly identify an icon without its name being present. You will see an icon for two seconds and then it will disappear. At that point you will have four seconds to respond. To respond press the appropriate

Function key (F1-F9) on the keyboard. If you do not respond within four seconds your answer will be considered incorrect. For this task you will be required to correctly identify all nine icons consecutively twice. Please notify the experimenter once this task has been completed.

Next you will be asked to participate in the actual test. To begin, click on the "Start" button on the screen. You will be asked to view 162 3x3 displays of nine aircraft icons. You will see the display for seven seconds. After seven seconds the display will disappear and be replaced with a question and response screen. On this screen you will be asked to identify an aircraft icon that was in one of the nine positions in the display you had just viewed. To respond to the question, press the appropriate Function key (F1-F9) on the keyboard. You may take as much time as needed to respond. Please notify the experimenter once this task has been completed.

Once you have viewed all 162 displays you will be asked to complete the NASA TLX workload assessment scale. Please follow the NASA TLX instructions. Please notify the experimenter once this task has been completed.

If this is your first day of testing you have completed today's tasks. Please check with the experimenter to confirm your time and date for the second day of testing. Once you have completed your second day of testing you will be asked to complete a short post-test questionnaire. Please notify the experimenter once you have completed the post-test questionnaire.

INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974

CONSENT FORM

NEW TECHNOLOGIES FOR MAINTENANCE AND LOGISTICS INFORMATION SYSTEM STUDIES

1. Nature and Purpose: I have been invited to participate in research studies to evaluate new technology applications to the maintenance and/or logistics planning environments. The purpose of these studies is to evaluate such factors as data recall techniques, formats, and demonstration systems for presenting technical information prior to their incorporation into test systems. Field tests will be used to evaluate demonstration systems developed using Laboratory developed techniques, software, and hardware. The studies will be designed to evaluate the various techniques and demonstration systems in terms of their ability to effectively provide the user with the required information, acceptability to the user, and ability to support the mission of the maintenance and logistics organizations. All tests will be conducted at Air Force Bases. Anticipated number of subjects is 20.

2. Experimental Procedures: My participation in this study will require me to view a computer screen. During two separate sessions participants will view a 3X3 display of nine aircraft icons. After viewing the display the participants will be asked to correctly identify one of the icons. During one session all of the icons will be non-animated, and during the other session four of the icons will be animated and five will be non-animated. In each session participants will complete 162 trials. Each session should last for about 1-1.5 hours.

3. Discomfort and Risks: I understand that I cannot participate in the experiment if I suffer from any known adverse effects from flashing lights (i.e. photosensitive epilepsy). My participation will not involve risks greater than I encounter performing my normal duties. I understand that depending on the length of time of testing I may experience eye fatigue from viewing so many displays.

4. Precautions for Female Subjects: There are no special precautions for female subjects.

5. Benefits:

- There I will not receive any known medial benefits resulting from participation in this experiment.

- My participation in this study will help to ensure that the application and further development of these technologies are designed to meet my needs. The ultimate benefit of this project will be to make maintenance and logistics personnel more effective and make their jobs easier. The only other way to obtain the required information would be to conduct studies in a laboratory setting using non-maintenance personnel. These people would not be representative of maintenance personnel, and the information gathered would not reflect the true needs of maintenance personnel. I am encouraged to provide the experimenter with feedback about the experiment so that my concerns can be considered in future investigations.

6. Entitlements and Confidentiality: Records of my participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 5 U.S.C. 552a, and its implementing regulations.

I understand my entitlements to medical and dental care and/or compensation in the event of injury are governed by federal laws and regulations, and that if I desire further information I may contact the base legal office (88ABW/JA - phone 257-6142).

If an unanticipated event (medical misadventure) occurs during my participation in this study, I will be informed. If I am not competent at the time to understand the nature of the event, such information will be brought to the attention of my next of kin.

The decision to participate in this research is completely voluntary on my part. No one has coerced or intimidated me into participating in this program. I am participating because I want to. Mr. Carlton Donahoo. Mr. Donahoo has adequately answered any and all questions I have about this study, my participation, and the procedures involved. I understand that Mr. Donahoo, or her representative; will be available to answer any questions concerning procedures throughout this study. I understand that if significant new findings develop during the course of this research, which may relate to my decision to continue participation, I will be informed. I further understand that I may withdraw this consent at any time and discontinue further participation in this study without prejudice to my entitlements. I also understand that the medical monitor of this study may terminate my participation in this study if she or he feels this to be in my best interest.

7. Alternatives: Choosing not to participate is an alternative to participating in this study.

VOLUNTEER SIGNATURE AND SSAN (Optional) DATE and TIME

PRINCIPAL INVESTIGATOR SIGNATURE DATE

WITNESS SIGNATURE DATE

Privacy Act Statement

Authority: We are requesting disclosure of personal information, to include your Social Security Number. Researchers are authorized to collect personal information (including social security numbers) on research subjects under The Privacy Act-5 USC 552a, 10 USC 55, 10 USC 8013, 32 CFR part 219, 45 CFR Part 46, and EO 9397, November 1943 (SSN).

Purpose: It is possible that latent risks or injuries inherent in this experiment will not be discovered until some time in the future. The purpose of collecting this information is to aid researchers in locating you at a future date if further disclosures are appropriate.

Routine Uses: Information (including name and SSN) may be furnished to Federal, State and local agencies for any uses published by the Air Force in the Federal Register, 52 FR 16431, to include, furtherance of the research involved with this study and to provide medical care.

Disclosure: Disclosure of the requested information is voluntary. No adverse action whatsoever will be taken against you, and no privilege will be denied you based on the fact you do not disclose this information. However, your participation in this study may be impacted by a refusal to provide this information.

DEBRIEFING FORM – Computer Icon Animation

Thank you very much for participating in this study. I would now like to address any questions that you may have, and to give you some information concerning the purpose of the study. The main purpose of this study was to determine if animated icons better conveyed their meaning than non-animated icons, thereby making them easier to identify. The animation was incorporated into some of the icons in order to make them either more isomorphic to their physical counterparts, or to make them more distinct (stand out) from the surrounding icons. An icon is considered to be isomorphic when it resembles its physical counterpart with as little ambiguity as possible. By making the icon similar to its physical counterpart or making it more distinctive, it may be more easily identified.

Another goal of the study was to determine the ease/difficulty in identifying a target within a display composed of both animated and non-animated icons, and in identifying a target within a display composed of only non-animated icons. A third goal of the study was to determine if the amount of animation within a display had any effect on the identification of a target. For further information on this psychological research you may review the articles below.

References

Alpert, S.R. (1991). Self-describing animated icons for human-computer interaction: A research note. *Behaviour and Information Technology,* 10, 2, 149-152.

Bonder, R.C., & MacKenzie, I.S. (1997). Using animated icons to present complex tasks. *Proceedings of CASCON'97,* 281-291.

Rogers, Y. (1989). Icons at the interface: Their usefulness. *Interacting with Computers,* 1, 1, 105-117.

Assurance of Privacy

We are seeking general principles of behavior and are not evaluating you personally in any way. Your responses will be confidential and your responses will only be identified by a participant number in the data set along with other participants' numbers.

Contact Information

Participants may contact Carlton Donahoo, WPAFB, building190, 937.256.9243 if you have questions or problems after the study. Participants may also contact the chair of the Research Review and Ethics Committee, Dr. Charles Kimble at the University of Dayton, St. Joes Hall, room 319 or call 937.229.2167. Thank you for your participation.

Here is the information that you may need about this study. You may take this page with you.

References

Alpert, S.R. (1991). Self-describing animated icons for human-computer interaction: A research note. *Behaviour and Information Technology,* 10, 2, 149-152.

Bonder, R.C., & MacKenzie, I.S. (1997). Using animated icons to present complex tasks. *Proceedings of CASCON'97,* 281-291.

Rogers, Y. (1989). Icons at the interface: Their usefulness. *Interacting with Computers,* 1, 1, 105-117.

Contact Information

Participants may contact Carlton Donahoo, WPAFB, building190, 937.256.9243, if you have questions or problems after the study. Participants may also contact the chair of the Research Review and Ethics Committee, Dr. Charles Kimble in SJ 319, 937.229.2167.

APPENDIX B

NASA TLX Workload Measures

SUBJECT INSTURCTIONS: RATING SCALES

I am not only interested in assessing your performance but also the experiences you had during the different task conditions. Right now I am going to describe the technique that will be used to examine your experiences. In the most general sense I am examining the "workload" you experienced. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The workload contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors I would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

After performing each of the tasks, you will be given a sheet of rating scales. You will evaluate the task by putting an "X" on each of the six scales at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "own performance" goes from "good" on the left to "bad" on the right. This order has been confusing for some people. Please consider you responses carefully in distinguishing among the different task conditions. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment and is greatly appreciated.

SUBJECT INSTRUCTIONS: SOURCES-OF-WORKLOAD EVALUATION

Throughout this experiment the rating scales are used to assess your experiences in the different task conditions. Scales of this sort are extremely useful, but their utility suffers from the tendency people have to interpret them in individual ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended on a given task or the level performance they achieved. Others feel that if they performed well the workload must have been low and if they performed badly it must have been high. Yet others feel that effort or feelings of frustration are the most important factors in workload: and so on. The results of previous studies have already found every conceivable pattern of values. In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort vs. Mental Demands) and asked to choose which of the items was more important to your experience of workload in the task(s) that you just performed. Each pair of scale titles will appear on a separate card.

Circle the Scale Title that represents the more important contributor to the workload for the specific task(s) you performed in this experiment.

After you have finished the entire series I will be able to use the pattern of your choices to create a weighted combination of the ratings from that task into a summary workload score. Please consider your choices carefully and make them consistent with how you used the rating scales during the particular task you were asked to evaluate. Don't think that there is any *correct* pattern: I am only interested in your opinions.

If you have any questions, please ask then now. Otherwise, start whenever you are ready. Thank you for you participation.

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, looking, searching. Etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHSYICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Subject ID:________________ Task ID:________________

RATING SHEET

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

Effort Or Performance	Temporal Demand Or Frustration
Temporal Demand Or Effort	Physical Demand Or Frustration
Performance Or Frustration	Physical Demand Or Temporal Demand
Physical Demand Or Performance	Temporal Demand Or Mental Demand

Frustration
Or
Effort

Performance
Or
Mental Demand

Performance
Or
Temporal Demand

Mental Demand
Or
Effort

Mental Demand
Or
Physical Demand

Effort
Or
Physical Demand

Frustration
Or
Mental Demand

APPENDIX C

Preference Questionnaire

Please answer the following questions regarding your experiences with both the animated and non-animated icon formats.

(1) The Taxi icon was presented in both the animated and the non-animated formats. **Please place a check by the format that you preferred.**

____ Animated icons
____ Non-animated icons

Please explain why.

(2) The Engine Start icon was presented in both the animated and the non-animated formats. **Please place a check by the format that you preferred.**

____ Animated icons
____ Non-animated icons

Please explain why.

(3) The Pre-Flight icon was presented in both the animated and the non-animated formats. **Please place a check by the format that you preferred.**

____ Animated icons
____ Non-animated icons

Please explain why.

(4) The In Flight icon was presented in both the animated and the non-animated formats. **Please place a check by the format that you preferred.**

____ Animated icons
____ Non-animated icons

Please explain why.