A R L

Aviation Research Lab
Institute of Aviation

University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874

ATTENTION AND TRUST BIASES IN THE DESIGN OF AUGMENTED REALITY DISPLAYS

Michelle Yeh and
Christopher D. Wickens

Technical Report ARL-00-3/FED-LAB-00-1

April 2000

Prepared for

U.S. Army Research Laboratory
Interactive Displays Federated Laboratory

Contract DAAL 01-96-2-0003

ABSTRACT

This experiment seeks to examine the relationship between three advanced technology features (presentation of target cueing - and the reliability of that data, image reality, and interactivity) and the attention and trust provided to that information. In particular, we investigate the nature of two sorts of biases: an attention bias, in which the operator focuses attention to an area highlighted by the automation at the expense of other areas of the visual scene, and a trust bias, in which unwarranted attention is given to the guidance information.

Sixteen military personnel searched for targets camouflaged in terrain, presented at two levels of scene detail (manipulated by varying the number of polygons with which the scene was generated and the level of the detail with which the scene was textured) while performing a terrain association task. Half the subjects actively navigated through the terrain; the other half passively viewed the control path of an active navigator. Cueing was presented for some of the targets, and the reliability of this information was manipulated at two levels (100% and 75%). More importantly, there were objects in the environment that were of high priority and uncued, and detection of these targets provided an assessment as to the degree of an operator's reliance on the cueing information. To assess trust in the terrain simulation, subjects were asked to perform terrain association given a hand-held paper map and report any inconsistencies between the visualization and the map.

The results showed that the presence of cueing aided the target detection task for expected targets but drew attention away from the presence of unexpected targets in the environment. Cueing benefits and attentional tunneling were both reduced when cueing became less reliable. Unfortunately, despite knowing that the cue was only partially reliable, subjects still trusted the cue more than was warranted as evidenced by an increased false alarm rate, a drop in sensitivity, and a risky criterion shift. The degree of this change in sensitivity was so great that the loss of reliability in the cue eliminated the benefits of cueing.

Finally, neither image reality nor interactivity directly influenced trust in the simulation itself. Instead, the influences of interactivity were attributable to increased resource demand, which modulated performance in the presentation of unreliable cueing. Enhancing the image realism of the scene was compelling and increased reliance on the cueing information when that data was reliable.

TABLE OF CONTENTS

ABSTRACT 1
1. INTRODUCTION 3
2. ATTENTIONAL BIASES 4
2.1 Superimposing Imagery 5
2.2 Cueing Effects: Directing Attention 7
2.3 Frame of Reference (FOR) 8
2.4 Summary 10
3. AUTOMATION RELIABILITY: TRUST BIASES 11
3.1 Automation Overtrust 12
3.2 Automation Undertrust 13
3.3 Summary 14
4. ATTENTION AND TRUST BIASES TOGETHER 14
4.1 Increasing Scene Realism 16
4.2 Interactivity 20
5. RESEARCH QUESTION 23
6. METHOD 25
6.1 Subjects 25
6.2 Task Overview 25
6.3 Apparatus 26
6.4 Displays/Tasks 26
6.5 Experiment Design 29
6.6 Procedure 30
6.7 Performance Measures 31
7. RESULTS 32
Overview 32
7.1 Target Detection 32
7.2 Image Realism 34
7.3 Interactivity 35
7.4 Trust Effects and Biases: Cue Reliability 37
7.5 Trust Effects due to Image Realism: Checkpoint Detection 45
7.6 Subjective Measures 46
8. DISCUSSION 48
8.1 Overview 48
8.2 Cueing 49
8.3 Reliability 50
8.4 Interactivity 51
8.5 Image Realism 52
8.6 Model of Attention and Trust Biases 54
9. CONCLUSIONS 55
9.1 Theoretical Implications 55
9.2 Design Implications 56
ACKNOWLEDGEMENTS 57
REFERENCES 58
APPENDIX 1: STUDIES OF TRUST AND ATTENTIONAL BIASES 69
APPENDIX 2: PILOT STUDY 78
APPENDIX 3: EXPERIMENT INSTRUCTIONS 81
APPENDIX 4: QUESTIONNAIRE 84
APPENDIX 5: ANOVA TABLES 89

1. INTRODUCTION

Computer generated imagery is an invaluable tool for developing visual models of terrain data in order to obtain navigational information and even provides a means for scientists to visualize and test theories and hypotheses of abstract concepts. The need to present such complex information has led to display enhancements which allow a more efficient presentation of data through a more “invisible” interface using techniques of **augmented reality**, in which supplementary information relevant to the task at hand is referenced to the “real” world beyond. This “real” world can be any representation of an environment whether it is created from sampled image data (e.g., video, photographs, radar, x-rays, and ultrasound) or not. An example of an augmented reality is a scene developed from raw sensor data onto which textured polygons are mapped to create a “real” world view, and such an image may be augmented by cueing symbology which calls attention to interesting aspects of the visual scene (Drascic & Milgram, 1996; Milgram & Colquhoun, 1999).

Much of the interest in the development of augmented reality displays is focused on how they can be used to aid the operator, who in trying to determine the status of a situation, must collect and integrate information from various sources in order to select an action. Consider, for example, the task faced by the military commander who may be asked to allocate attention to and integrate information regarding enemy and friendly unit positions and movement in order to make decisions about a future course of action or the pilot trying to locate his position while examining information on a digital map and searching for corresponding features in the environment.

To support users’ interactions in these complex spatial environments, several advanced display technology features have been proposed: superimposing imagery in the forward field of view, e.g., attached to the windshield of an aircraft or a car using a head-up display (HUD) or through a wearable computer such as a helmet-mounted display (HMD); displaying the information using an immersed frame of reference; presenting intelligent cueing information to guide attention throughout the scene; and increasing the image reality and interactivity. However, each of these display features is costly in terms of implementation, and more importantly, each of these display features is *compelling*, a characteristic which may influence what information is noticed by the user as well as the level of confidence that users attribute to the validity of that information *whether the display designer intended it to be noticed or not*. Consequently, the incorporation of these augmented reality display features may induce two important costs to determining the true reliability of an information source. The first is a bias, or miscalibration, in ***attention***, such that attention is inappropriately allocated between two or more sources given the value of the data provided or is inappropriately focused on one specific source of information, e.g., the operator who fails to monitor the rest of the visual scene for objects of interest, i.e., attentional tunneling (Mosier & Skitka, 1996). When guidance information (e.g., a cue indicating the location of a target) is provided, an attentional bias favoring the areas highlighted by the guidance symbology results. If the guidance information is unreliable, then a second bias is produced – a miscalibration in ***trust***. This is evidenced by two types of behavior: operators may exhibit signs of *overtrust* in the information source and rely on the data provided, failing to seek out more information when it is critical to do so, or *undertrust*, in which operators assume the guidance information is incorrect – because it has been wrong before, and fail to rely on it when it is appropriate to do so (Wickens, Pringle, & Merlo, 1999).

In the following pages, we consider display features which have been proposed for augmented reality technology in order to determine (1) the performance benefits each feature offers, and (2) more importantly, whether some (or all) impose potential performance costs associated with attention and trust biases. Figure 1.1 links these display features with their potential costs in inducing departures from optimal calibration of attention and trust.

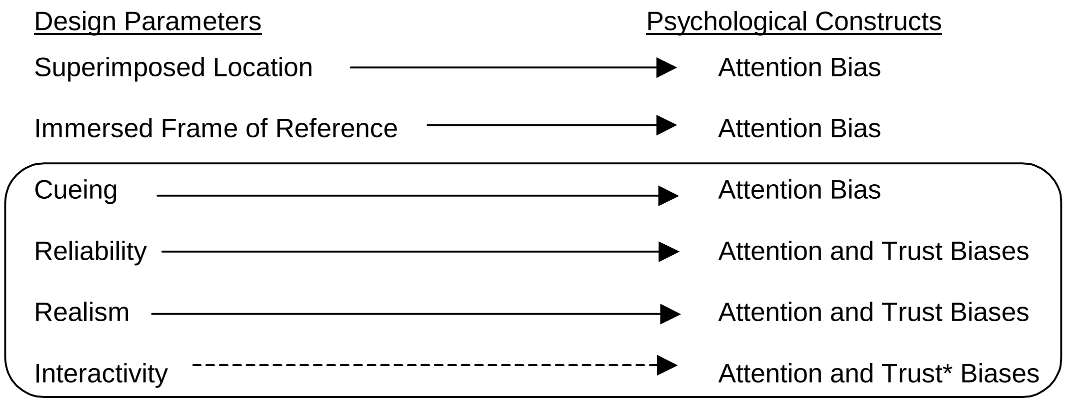


Figure 1.1. Biases mediated by augmented reality display features.
(* In this case, the trust bias is expected to be an *overreliance* as the amount of interactivity increases.)

In Figure 1.1, the solid arrows describe those effects that *directly* influence calibrations of attention and/or trust; the dotted arrow describes that effect which are mediated by task demands, and *indirectly* influence attention allocation and trust. The variables presented within the box are the ones we propose to manipulate in the current study.

Evidence will be presented to indicate that calibrations in attention are mediated by task and display characteristics. Biases in this calibration have been found to be a consequence of superimposing two sources of information in one location, presenting an immersed frame of reference, and presenting highly accurate guidance (cueing) information. Additionally, we compare behavior when the guidance information is 100% reliable to when it is only partially reliable, and in this context, discuss how attention allocation changes as a consequence of trust in the cue, mediated by its reliability or accuracy. In particular, the review focuses on how instances of overtrust in automation influence attention allocation, i.e., when operators attend to the information highlighted by the guidance information to such a degree that attention is less focused on the visual scene.

We then turn to explore the use of two display enhancements – image reality and interactivity, which have received a great amount of attention by computer scientists, engineers, and psychologists in the design of simulators and virtual environments given that very little is known about what actual benefits it provides (Salas, Bowers, & Rhodenizer, 1998). The literature is less clear in describing how these features will affect trust and therefore attentional or trust biases, but there is reason and concern to believe that such effects will be present (Theunissen, 1998). In considering the effects of enhancing image reality, a reasonable hypothesis to be investigated is that any guidance cue associated with a more realistic scene may be trusted more and hence, receive more attention at the expense of attention to the properties of the scene itself. Similarly, the literature provides no evidence that increased interactivity will influence trust, despite the fact that interactivity is more “real”. However, interactivity is a feature being proposed for new applications and is a feature that influences attention allocation because of its tasks demands (e.g., interactivity forces the user’s attention to be allocated to guidance cues). Furthermore, because of the increased workload associated with interactivity, it could well increase the dependency upon automated cueing, and hence the amount of attention to what is cued, even if it does not increase trust in the cue. Note that summaries of the studies can be found in Appendix 1, which describes the characteristics of the experiment and its evidence for an attention and/or trust bias.

2. ATTENTIONAL BIASES

The presentation of display features which are compelling influences the operator’s calibration of attention which should be allocated to an information source. Features which command attention such as color or intensity, naturally draw one’s attention, and when those features are applied to target objects in a display, i.e., those objects that represent the most important information, the time to find the necessary data is reduced (Fisher, Coury, Tengs, & Duffy, 1989; Fisher & Tan, 1989; Nagy & Sanchez, 1992). Additionally, the presentation of such cues are often associated as representing characteristics of importance and reliability (Berg, 1990; Montgomery & Sorkin, 1996). In the design of augmented reality displays, there is evidence that the compellingness of an information source can be induced by: (1) superimposing information at the same location (Fadden, Ververs, & Wickens, 1998; McCann, Foyle, & Johnston, 1992; National Research Council, 1997; Wickens & Long, 1995), (2) presenting highly accurate attentional guidance (cueing) information (Yeh, Wickens, & Seagull, 1998; Ockerman & Pritchett, 1998), and (3) increasing the apparent realism of the imagery by presenting information using an immersed perspective (Olmos, Wickens, & Chudy, 1997; Wickens et al., 1999). We consider each of these potential sources of increased compellingness below and examine their effects on the calibration of attention.

2.1 Superimposing Imagery

The use of a see-through HMD presents the opportunity for superimposing images from the near domain onto the far domain (the real world), allowing the user to perform tasks with a head-mounted "guide" with the benefit of maintaining the high level of detail present in the real world while reducing the amount of computational resources which would otherwise be required to update computer-generated scenes of a complex outside world (Barfield, Rosenberg, & Lotens, 1995). However, such presentation imposes new attentional demands and requires that the information in the two domains (the near domain imagery and the far domain world) be integrated in such a way that the viewer considers them as one. In doing so, the configuration of the display must lend itself to aid the completion tasks of both focused attention in one domain – and thus filtering out irrelevant information in the second domain – and divided attention – e.g. tracking data in the near domain while monitoring for hazards in the world.

2.1.1 Focused Attention

The resulting visual clutter from two overlapping sources of information has been shown to interfere with tasks of focused attention as superimposing information on the forward field of view may obscure events in the outside world. In a flight simulation study, Ververs and Wickens (1998b) found that focusing attention in the near domain was difficult due to the low contrast between the HUD symbology and outside world in a task which required subjects to focus on flightpath information in the near domain and scan for aircraft in the far domain. In fact, monitoring information on the HUD was impaired by the presence of objects in the outside world.

To aid the focused attention task, the basic attentional literature shows that attention may be allocated to particular salient display features, e.g., color, highlighting, flashing, or size. For example, the use of color-coding allows stimuli presented in the target color to "pop out" so that search proceeds in parallel. In this case, an increase in the number of non-target items only has a slight increase on the search time – or none at all since items that differ in color from the target can be rejected rapidly (Treisman, 1988). An alternative feature is highlighting or intensity coding, in which items important to the task at hand are presented in a higher intensity than those items which are currently inconsequential. This technique has been proven to be effective in calling attention to desired information whether on a computer display (Fisher & Tan, 1989; Fisher, Coury, Tengs, & Duffy, 1989) or on a HUD (Ververs & Wickens, 1998b).

However, when these compelling display attributes are applied to *irrelevant* items, a consequence of this attentional guidance may be more processing of the "highlighted" feature and less processing of what may really be important in the display. In the auditory domain, Berg (1990) finds that observers assume that the most intense tones are the most reliable, and these tones often are given the most weight, even when they are the least reliable, i.e., when high intensity represents low reliability. In fact, listeners may implicitly assign a sense of urgency to an auditory cue, and for the designer who is not careful in his choice of auditory alerts, a mismatch may occur between the perceived urgency and the actual urgency of the situation (Edworthy, Loxley, & Dennis, 1991).

2.1.2 Divided Attention

The ability to present information superimposed on the user's forward field of view, e.g., the presentation of guidance symbology, reduces the amount of time the operator would have spent head down accessing this information, e.g., from the dashboard of a car or a hand-held display, and if the display is collimated, reduces the amount of eye accommodation required for switching focus between the near domain (symbology) and far domain (world). In the field of medical imaging, doctors can view a virtual image of their patient's internal organs superimposed on the patient's body using an HMD (Rolland, Biocca, Barlow, & Kancerla, 1995) and in the field of aircraft inspection, HMDs provide hands free operation for a mobile operator performing tasks requiring high information content (Ockerman & Pritchett, 1998).

However, these benefits must be weighted against the potential costs to superimposing symbology, and more specifically, doing so using a see-through HMD. Information presented in the near domain (i.e., the display) clutters the forward field of view and, in a worst case scenario, may obscure information in the real world (or far domain). Additionally, people often have problems treating two overlapping sources of information as one. When Neisser and Becklen (1975) showed subjects superimposed videos of people playing games and asked them to follow the action in one or both games, they found that subjects could easily focus their attention on only one game but were unable to pay attention to both games without practice. In fact, subjects were so focused on one game, that often they did not notice the figures in the other game performing actions which were unexpected and out of context. Becklen and Cervone (1984) re-investigated this issue and agreed that effectively dividing attention between the two games required extensive practice.

The difficulty dividing attention between two sources of information presented by overlapping imagery has been examined extensively in the HUD and HMD domains and is represented in the cases in which the symbology presented in the forward field of view captures the pilots' attention at the cost of processing other information in the far domain beyond the symbology, i.e., *attentional tunneling* (Fadden et al., 1998; Larish & Wickens, 1991; Martin-Emerson & Wickens, 1992, 1997; McCann et al., 1992; National Research Council, 1997; Sampson, 1993; Wickens & Long, 1995). This miscalibration in attention may be a consequence of emphasizing certain display features [e.g., superimposing guidance symbology as in Ververs and Wickens (1998a)] such that operators overutilize the information provided by the system, which allows performance of the task with less cognitive effort. In the case of displaying imagery on a HUD or HMD, the presentation of information superimposed on the forward field of view in the design of augmented reality displays is more compelling than information presented in a head-down location where it appears less "natural" and direct and requires a greater degree of cognitive transformation to identify a direct position in space. In fact, Dopping-Hepenstal (1981) noted that military pilots fixated more frequently on information presented on a HUD at the cost of scanning the outside scene but at the same time were more confident in their abilities. Fadden et al. (1998) conducted a meta-analysis of research that has compared the use of head-up versus head-down displays in air and ground vehicles and found that as more information is added to the HUD, there is a cost in detecting events in the world (e.g., noticing the presence of an airplane on the runway during a landing task) relative to the head-down display, particularly if these events are unexpected and not salient (Larish & Wickens, 1991; Wickens & Long, 1995; Wickens, 1997).

In one such study, Wickens and Long (1995) asked pilots to fly along a pre-specified flight path maintaining vertical and lateral tracking before breaking out of the clouds, seeing the runway, and executing a landing task. The symbology used consisted of a virtual runway or non-conformal instrument landing system symbology and was presented head-up or head-down. On the very last landing trial of the experiment, pilots were presented with a simulated runway incursion and were asked to report the presence of a plane on the runway and initiate a go-around. The results showed that pilots failed to notice an airplane on the runway more often when using the head-up versus the head-down display.

Similar results have been found in the HMD domain. Yeh and Wickens (1998) found that subjects were more likely to constrain their search field when information was presented on a conformal HMD than a hand-held display. Subjects were presented with cueing information which was presented conformally in a head-up location or non-conformally head-down on a hand-held display. The results showed that the conformal cueing information presented head-up was more compelling than when presented head down causing subjects to attend to and rely too much on the guidance information and miss the presence of the unexpected high priority target. This effect was mediated by the display location, such that when cueing information was presented in a less real format on a hand-held display, attentional tunneling decreased.

Other data suggests that the costs of attentional tunneling for superimposing symbology may be reduced when the imagery in the near domain is "scene-linked" to the objects in the real world. In this type of display, a virtual copy of the information in the real world is presented in the forward field of view such that information in the near domain is “linked” to objects in the far domain enabling the user to collect and combine natural cues present in the environment in order to complete his task (McCann & Foyle, 1995). In an aviation flight simulation, Foyle, McCann, and Shelden (1995) and Levy, Foyle, and McCann (1998) found that dividing attention between the symbology and the real world was not hindered if symbology was presented conformally in such a way that the two domains could be “fused” to better allow the user to treat both domains as one. Subjects were asked to maintain their altitude and follow a ground path in conditions in which altitude information was superimposed either non-conformally or conformally along the flight path. The availability of non-conformal flight information improved performance on the altitude maintenance task but resulted in poorer performance following the ground path. However, this trade-off was not present for the conformal displays, in which the altitude information on the display was linked to the lateral guidance cues in the real world.

Similarly, in the HUD landing experiment described earlier, Wickens and Long (1995) found that focused attention on the far domain to detect runway incursions benefited from the presentation of conformal scene-linked symbology on the HUD. The authors attributed this result to the fact that superimposing imagery (e.g., a runway outline) with its far domain counterpart (e.g., the true runway) created a sense of “fusion” between the near and far domain (the symbology and the real world, respectively), hence possibly “scene linking” the two (Foyle, Andre, McCann, Wenzel, Begault, & Battiste, 1996) in a way that would benefit divided attention to both. Finally, Ververs and Wickens (1998a) noted greater evidence for attentional tunneling (failing to notice an unexpected aircraft) in a HUD, when the HUD did not provide fused, scene-linked symbology with the far domain, than in a condition in which such scene linking was present (ground taxiing).

When a cue (e.g., an arrow pointing the location of a target) is used to orient attention, operators may inappropriately allocate attention to the cued location and fail to attend to sources of information that while uncued, are still relevant (Ockerman & Pritchett, 1998; Yeh, Wickens, & Seagull, 1998). This is a display enhancement we now examine in greater detail.

2.2 Cueing Effects: Directing Attention

It is relatively straightforward to predict that cueing a target location will facilitate the detection of that target (Ball & Sekular, 1981; Flannagan et al., 1998; Posner, 1978; Posner & Snyder, 1974). Furthermore, extrapolating from basic research, one can predict that directly overlapping the cue on a target will provide faster and more accurate cueing than a less direct means of guiding attention – e.g., an arrow pointing to the cue (Egeth & Yantis, 1997; Jonides, 1981; Müller & Rabbitt, 1989). However, on an HMD, the ability to provide such highly precise cueing information is subject to factors such as time delay, incorrect reports from intelligence analysts or sensors, and errors in head-tracking from automation aids, and this imprecision may make the cueing data less informative. It is easy to imagine that a cue that does not overlap the target will still produce performance benefits if it is relatively close to the target. On the other hand, a very imprecise cue may redirect one’s focus away from the object of interest, and such a cue is more costly (as measured by target detection time) than presenting no cueing information at all (Ball & Sekular, 1980; Hübner, 1996; Mohindra, Spencer, & Lambert, 1986; Posner, 1978; Posner, Snyder, & Davidson, 1980; Umiltá, Riggio, Dascola, & Rizzolatti, 1991).

The issue of how attention is guided through the display has been examined in the basic attentional literature by comparing the effectiveness of <u>central cueing</u>, in which a cue conveys information about target location in the center of the observer’s visual field regardless of the target’s real-world location, to <u>peripheral cueing</u>, in which cueing information is presented at the exact location where the target will appear. In these studies (e.g., Müller & Rabbitt, 1989; Rizzolatti, Riggio, Dascola, & Umiltá, 1987; Umiltá et al., 1991), subjects searched for a target, which was often a luminance change or the onset of an LED, and cueing information in the form of an arrow pointed to the target or a flash was presented at the target’s location. In general, the results showed that attention was allocated more effectively when subjects were presented with a peripheral cue than with a central cue, as the physical property of the former triggered a response by a fast acting reflexive (automatic) mechanism whereas the information presented by the latter first needed to be “decoded” in order to determine the spatial location designated. In fact, when attention was oriented with central cues to an area more than 4° away from the target, detection time slowed as a function of the distance from the cued location to the target location. On the other hand, no such cost was observed with peripheral cueing. In this case, the degree of registration accuracy between the cued location and the target location could be expanded to an area within 8° of the target. Thus, these findings suggest that the effects of cueing for areas beyond 4° of the target location are enhanced with cues located close to the targets, relative to less direct means of attentional guidance (e.g., central cueing).

The findings that cueing benefits are enhanced with cues located close to (or superimposed over) the targets, relative to less direct means of attentional guidance (Jonides, 1981), suggests that the cueing benefits will be greater with augmented reality imagery. For example, the closer the cue is to the target, the more the user may trust the information provided by the automation and the more attention will be allocated to the region around the cue, consequently reducing the allocation of attention to the rest of the visual scene. This hypothesis was supported by the results of Yeh, Wickens, and Seagull (1998), who asked Army cadets and officers to search for targets consisting of tanks, soldiers, land mines, and nuclear devices camouflaged within a terrain environment. Cueing information, in the form of an arrow which changed to a reticle box when the target was visible within the center 40° of the subject's field of view, was displayed on an HMD and superimposed over the terrain environment. The nuclear device served as the high priority target and was presented on only 10% of the trials; when it was present, however, subjects were instructed that detecting the nuclear device took priority over detection of any of the other targets. The results showed that the presence of cueing aided the target detection task for expected targets but drew attention away from the presence of unexpected targets in the environment: the high priority target was more likely to be overlooked if it appeared on the same trial as a cued tank than as an uncued soldier.

Ockerman and Pritchett (1998) found similar results in the context of an airline maintenance inspection task – when operators wore an HMD which presented data to assist in aircraft inspection, the operators were more likely to miss faults that were not cued. Subjects were asked to perform a pre-flight inspection with the aid of an HMD-based wearable computer. A list of potential aircraft faults was presented on the HMD in either a text or picture format. A control condition required the aircraft inspector to search for faults using his own memory. Subjects' fault detection performance was improved for those faults for which inspection guidance information was presented on the HMD but suffered for those faults which were not listed on the HMD, relative to the control condition. Furthermore, when a more "compelling" picture based computer image was employed, more non-cued failures were missed than when a text based image was employed. Thus, while explicitly guiding attention to certain key features of the display is somewhat beneficial (e.g., the presentation of a cue indicating the location of a target significantly reduces the visual search space), implicitly, the guidance information draws attention away from uncued objects hidden in the environment which may be just as important as the cued ones.

So far, studies of attention allocation from a psychological standpoint have not examined attention allocation purely as a function of trust and system reliability. While Posner and his colleagues have manipulated the reliability of cueing symbology and measured the attentional cost for doing so, these studies confound the measures of attention and trust by assessing the attentional cost given to an *unreliable* cue. That is, attentional tunneling was characterized as trusting the cue when it guided attention to an incorrect region in space. However, attentional tunneling as implemented in the automation paradigms studied by Merlo, Wickens, and Yeh (1999), Ockerman and Pritchett (1998), Yeh, Wickens, and Seagull (1999) has been assessed not by attentional behavior when the cue was incorrect, but rather by introducing the idea of *importance* and examining reliance on cueing information when the cue was *correct* but simply did not guide attention to the most critical event in the scene.

This attentional bias may be attributed to "cognitive laziness," an inclination to perform tasks using as little cognitive effort as possible (Mosier, Skitka, Heers, & Burdick, 1998). For example, when making decisions, one often applies simplifying heuristics to reduce the amount of information that needs to be considered at once (Wickens, 1992). Further evidence for cognitive laziness is evidenced by examining user interaction with information presented in an immersed frame of reference.

2.3 Frame of Reference (FOR)

In an augmented reality environment, the view with which the user is presented with may be *egocentric* such that the user is immersed in the display and views and interacts with the real or virtual world from his own perspective, or *exocentric*, which provides a view from a point fixed at some height above the user's current position, such that the user's movement could be tracked but not his orientation (Wickens & Baker, 1995). Note that the level of egocentricity falls along a continuum and can be varied in depicting the lateral or vertical dimensions creating a *tethered* view; in this case, the viewpoint is from a position behind one's actual position in the space creating a wider view of the environment than with the egocentric view but not to the same extent as the exocentric view (McCormick, Wickens, Banks, & Yeh, 1998; Wickens, 1999). It is somewhat intuitive to conclude that the view provided by an egocentric display is more realistic than the view from an exocentric display. In fact, Slater and Wilbur (1997) find that a system with which the user is able to interact through an immersed egocentric viewpoint is likely to improve the sense of "presence" or "the sense of being there" within a virtual environment.

The FOR with which information is presented influences what objects users attend to, and consequently, what is learned (Presson, DeLange, & Hazelrigg, 1989; Thorndyke & Hayes-Roth, 1982). The different FORs are suited for different types of tasks. In search, the user can locate items faster if he can view the entire information space at once; this need is met by the exocentric FOR, which provides a wide field of view which allows the user to examine the entire virtual environment at once without the need to change orientation. On the other hand, the perspective with which the user interacts with the environment using the egocentric FOR is in a sense limited to a "key hole" since the user can only search what is directly in front of him; movement with a control device (or turning of the head with head-slaved HMDs) would be required to search other areas of the display space not currently in the forward field of view (McCormick et al., 1998; Vincow & Wickens, 1998).

However, when the goal of interacting with the display is navigation, travel with the immersed viewpoint provides a natural compatibility of perspective and control such that the display viewpoint is identical to the axis of control rotation. This makes sense from an attentional standpoint since, when travelling, it is the region in front of the traveler that is of greatest importance and should therefore receive the benefit of attention. On the other hand, navigation using the exocentric view is inefficient due to the cognitive transformations which may be required between the display and the axis of control. That is, the use of a static birds-eye viewpoint with which the user views the entire scene implies that user's perspective will sometimes be opposite the axis of control such that movement would require a control reversal (Wickens, 1999).

Finally, global awareness of the information space benefits from the exocentric viewpoint (Arthur, Hancock, & Chrysler, 1997; Barfield, Rosenberg, & Furness, 1995; McCormick et al., 1998; Wickens & Prevett, 1995; Wickens et al., 1999). This result is not surprising since viewing the immersed display would require mentally piecing together various "snapshots" of the environment taken from different perspectives of virtual space to form a "big picture" whereas the exocentric display would present one global view presented from one angle. Ellis, Tharp, Grunwald, and Smith (1991) hypothesized that the use of exocentric displays could prevent misjudgment of viewing location or direction since the viewpoint would be from a "bird's eye" view of the world.

Implicitly, the display viewpoint influences the user's actions by biasing judgements as to what elements are important in the display (Ellis, McGreevy, & Hitchcock, 1987; Theunissen, 1998). For example, in the experiment conducted by Yeh and Wickens (1998) discussed earlier, superimposing conformal cueing symbology using an HMD over the far domain environment prevented subjects from performing as thorough a scan of the environment as was required to detect a high priority, unexpected target. This cost was mediated when the cueing information was presented in an exocentric format viewed on a hand-held display.

Evidence exists to suggest that users viewing an egocentric display assume that the information appearing directly in front in the forward field of view is more important than information that is located to the side or behind the current position (which is not viewable without turning the head) regardless of whether this data is critical to the task. Consequently, a sort of attentional tunneling may result. Olmos, Wickens, and Chudy (1997) found that pilots inappropriately allocated visual attention to a more compelling egocentric display at the cost of detecting events available on a concurrently viewable exocentric display. Pilots were asked to fly to eight waypoints while monitoring for hazards. Pilots using the immersed display required more time to respond to respond to traffic threats appearing in the exocentric display but not immediately viewable in the immersed egocentric display relative to pilots who viewed the full information on a single exocentric display. In other words, attention was allocated inefficiently and more of the pilots' attention was given to information in the more compelling egocentric display at the cost of detecting hazards in the exocentric display.

Similar results were found by Wickens et al. (1999) in the domain of battlefield information visualization. Army officers were asked to count the number of enemy targets present in the terrain using an egocentric or exocentric map display. In the egocentric display, some of the enemy targets were viewable at the start of the trial, but subjects needed to change their viewpoint and pan the area to view all the targets. On the other hand, when using the exocentric view, all the enemy targets in the terrain were readily viewable. The results showed that performance accuracy on the task benefited from the presentation of data in the less realistic exocentric view. Officers using the immersed display often undercounted the number of enemy targets by focusing mainly on the information which was initially viewable in the forward field of view at the onset of the trial and undercounting the number of targets appearing to the side or behind the initial viewpoint which could be located through display panning, exhibiting a behavior characteristic of attentional tunneling. That is, the egocentric display implicitly suggested that the information viewable at the onset of the trial was more important than information not yet viewable, and officers consequently "anchored" their answers on what they could initially see.

Alternatively, the results provide evidence for "cognitive laziness," in which the FOR with which information is displayed is compelling not only because it conveys a greater degree of realism but also because it eases the operator's task (Mosier et al., 1998; Theunissen, 1998). In the study by Olmos et al., the egocentric display simplified travel through the environment relative to the exocentric display, and in the study by Wickens et al., officers may simply have failed to adequately pan the area due to the extra effort it required. Additional evidence suggests that when two sources of information are present (e.g., an egocentric view with an inset displaying an exocentric view), an attentional bias favors the display which simplifies the operator's task the most to such a degree that the user focuses on the information provided by that display. When that task is search or global awareness, the operator may vest himself in the information provided by a tethered or exocentric display (Pausch, Burnette, Brockway, & Weiblen, 1995; Vincow & Wickens, 1998).

As one example of this attention allocation bias, Stoakley, Conway, and Pausch (1995) developed the concept of a *world in miniature*, a global tethered view inset within a larger egocentric display. Interaction with the tethered view simplified interaction within the egocentric world by eliminating constraints inherent in the real world, allowing users to select and manipulate objects which would be out of reach or out of view in the larger immersed view. Although this inset was provided just as a viewing aid, Pausch et al. (1995) found that when an iconic view of the user was provided in this miniature world (a doll representation), users "cognitively vested" themselves in this tethered view, aligning their viewpoint in such a way that they could look "over the shoulder" of the doll icon, and used it as their primary source of information due to the increased efficiency of the interaction.

However, Vincow and Wickens (1998) find evidence that the attention allocation strategy that users may consider to be most efficient may not be the most optimal. Subjects were asked to navigate through a document information space and search for articles related to a given topic or keyword. The FOR with which subjects viewed the space was egocentric, exocentric, or egocentric with an exocentric map. The results showed that while performance was best with the exocentric display, subjects trusted the information in the exocentric view in miniature more than they should have.

Thus, the results from the FOR literature suggest that when users need to select between two sources of information, attention is allocated to the one which facilitates the user's task. When only a single source of information is provided by one source and that source is an immersed perspective, cognitive laziness manifests itself in a failure to carefully examine areas not visible in the forward field of view, resulting in a attentional tunneling on the information which is initially viewable (Wickens et al., 1999). The compellingness of the egocentric view may be enhanced by the increased resolution relative to the exocentric or tethered viewpoints, such that the elements in the display may look more real as they are depicted with a greater level of scene detail, an issue that will be addressed later.

2.4 Summary

The results of the literature provide evidence suggesting that some of the technologies proposed to enhance the user's interaction with a scene or the user's sense of reality in the scene may induce some biases in attention allocation. The presentation of cueing leads to biases in attention as shown by the unwarranted focusing of attention to the space around the cue (Ockerman & Pritchett, 1998; Yeh, Wickens, & Seagull, 1998). These focusing effects of cueing are enhanced by superimposing information which would have otherwise been presented in a head-down format onto the forward field of view using a HUD or HMD. In this case, attentional biases are evidenced by attentional tunneling on the compelling information provided by the cue, which directs attention to specific areas only, and does so at the cost of other information which may be present in the far domain environment (the real world).

In many cases, the misallocation of attention within a visual scene results from a bias in trust, in which belief that information from an attentional guidance system is accurate causes the operator to direct attention to features on a display or in the world which may or may not demand the level of attention they receive. When automated cueing information is provided and superimposed on the visual scene, such attention guidance may lower one's response criterion, resulting in not only an increased willingness to report a target at the cued location but also an increased tendency to report nontargets at that same location: a false alarm (Swennsen, Hessel & Herman, 1977; Conejo & Wickens, 1997). In other words, when an operator is presented with multiple sources of information, one may be favored at the cost of another. Consequently, trust in the information provided by one source and the belief that that information is reliable may lead the operator "down the garden path," guiding attention away from another information source that is just as or more reliable than the one the operator attends to. This is an issue we now consider.

3. AUTOMATION RELIABILITY: TRUST BIASES

The human operator is often confronted with an overabundance of data from several information sources and is required to develop an assessment of the situation to form a decision or select a course of action. It can be assumed that most sources of information are somewhat unreliable as a consequence of uncertainty inherent in the world. For example, in the case of a visual search task for targets hidden in the military battlefield, uncertainty may be characterized by a failure of sensor technology to detect the presence of all objects of interest on the battlefield, a failure of the system which interprets the information collected by the sensor technology, or simply a failure by the soldier to process the information presented to him. What is important here is the extent to which more reliable information is given more weight and is trusted more, i.e., how the operator *calibrates* his trust or belief relative to the reliability of the information provided (Wickens, Pringle, & Merlo, 1999).

Ultimately, one's trust in the information provided influences the calibration of attention: one bases one's calibration on the perceived reliability of an information source so that attention can be allocated in a more optimal fashion to the more reliable sources of information such that they receive more processing and more weight in diagnosis. Ideally, the calibration of trust evolves through a learning process whereby an operator assesses the reliability and dependability of an information source based on the accuracy and outcomes of its advice (Muir, 1987). However, numerous miscalibrations have been noted. Research in signal detection theory describes the occurrence of a "sluggish beta", in which the decision criterion for indicating whether or not a signal is present is not adjusted as much as is required by the changes in the probability of the signal. This is especially pronounced when people must take probabilities into account; in this case, people are conservative and overestimate the occurrence of rare events (Wickens, 1992), e.g., physicians often do not adjust their response criterion for diagnosis optimally based on the frequency of occurrence of the disease (Balla, 1980, 1982). Additionally, miscalibrations in trust have been noted in terms of weighting the reliability of information sources. In a study regarding trust in eyewitness testimony, Schum (1975) found that jurors often failed to weigh eyewitness testimony consistently with the reliability of a source, overestimating the importance of an unreliable source.

Trust calibration has been studied more extensively by examining human response to automation and the extent to which humans rely on the information provided by an automated system (Muir, 1987; Muir & Moray, 1996; Parasuraman & Riley, 1997; Singh, Molloy, & Parasuraman, 1993). In this context, trust may be calibrated as a function of the operator's *perception* of system reliability independent of overall system performance (Muir & Moray, 1996), the level of understanding of the automated system's behavior (Sarter & Woods, 1994), one's confidence in one's abilities to adequately perform the task (Lee & Moray, 1992), task difficulty (Liu, Fuld, & Wickens, 1995; McFadden, Giesbrecht, & Gula, 1998), and task complexity (Riley, 1996). While trust is sometimes measured explicitly [e.g., by Lee & Moray (1992) and Muir & Moray (1996)], more often, it may be inferred implicitly from other aspects of operator behavior or judgement in those instances when the automation is in error.

In this section, the literature review focuses on those studies which explicitly examined interaction with partially reliable automation in order to find evidence for biases in trust calibration. Here, trust can be assessed either by cases of operator *overtrust* in the automation, the extent to which an operator trusts the information provided by an automated system more than is warranted, or *undertrust* to the extent that an operator fails to invoke the aid of an automated system even at times when it can perform the task more accurately and efficiently. Each of these issues are considered as follows.

3.1 Automation Overtrust

The introduction of an automated system into any environment changes the nature of work such that humans process information provided by an automated aid in an unpredictable and sometimes non-optimal manner. Overtrust results when the operator assumes that the system will operate normally and does not notice nor prepares for those instances in which it fails. This case has been described as *automation-induced complacency*, a phenomenon in which operators become dependent on automation to inform them of problems within the system, behave as if the automation is failsafe, and consequently fail to intervene when it becomes necessary (Parasuraman & Riley, 1997). An analysis of Aviation Safety Reporting System (ASRS) incidents referencing automated aids showed that complacency or nonvigilance were often the cause of errors in missing events: once crew members entrusted a task to the automated system, they assumed – and trusted – the system to perform the task correctly to such a degree that the amount of monitoring that was then devoted to the system was not adequate to catch mistakes, inconsistencies, or failures (Mosier & Skitka, 1996). Similar results were found by McFadden et al. (1998) when subjects were asked to monitor a set of moving targets and update their position when new data were received about the positions of existing targets or potential new targets. Performance of this task was either manual, in which subjects themselves tracked and updated the locations of targets, or automated with a tracker system which could collect and record data for specific targets over time. The reliability of the automation's performance was defined in terms of the precision (distance from the last recorded target marker) with which the tracker searched for target signals, and this was varied at four levels. The results revealed that reliance on the automation increased as system reliability increased, and when the automation's performance was highly reliable, errors made by the automated tracker were often undetected.

Evidence suggests that the advice provided by automation is compelling, and in some cases, is so compelling that operators heed the recommendation without first confirming whether or not an error actually exists. Wiener (1980) reports that a crew of a DC-9, taking off from Denver in 1977, trusted the tactile and auditory alerts warning that a stall was imminent despite information to the contrary (normal airspeed and pitch attitude indications) and chose to abort the take off resulting in injuries to passengers and severe damage to the aircraft. Similarly, Mosier, Palmer, and Degani (1992) found that when pilots were asked to diagnose a *potential* engine failure and to do so using either a paper checklist or an electronic checklist in which the automation diagnosed as much of the system as possible, the pilots using the electronic checklist were more likely to shut down the engine as advised by the automated system rather than rely on the state information as presented by the system itself.

In another study of automated overtrust, Mosier et al. (1998) asked pilots to fly a simulated path using electronic checklists, which could sense the state of the system, and during the flight, presented pilots with a false automated warning indicating the presence of an engine fire (although no such fire was actually present). The results showed that while all the pilots detected the engine fire alarm and shut down the indicated engine, they did so without diagnosing the situation in detail or attending to other cues that informed them the engine was normal. In fact, the degree of overtrust in automated advice was so significant that operators reported seeing confirming cues which were not actually present. The pilots reported that a combination of multiple cues had been present to help them diagnose the fire; in other words, the pilots created "false memory" data and remembered a pattern of cues consistent with what they would have ordinarily seen had the engine actually been on fire.

Several factors may account for this bias in trust. Riley (1996) found evidence that familiarity with and frequency of automation use increased reliance on an automated system thereby inducing overtrust. Subjects in his study were asked to perform a character sorting task, indicating if a character was a letter or a number, in conjunction with a distractor task which required the subject to correct deviations of a marker from a pre-set location. An automated aid was provided which helped perform the categorization task; the accuracy of this system was set to be 90% or 50% (simulating failure). The results showed that subjects who were pilots and highly experienced with working with very advanced automated systems were more likely to invoke the use of an automated system – even during very high levels of automated failure – than subjects who were students and not as familiar with automation.

Not surprisingly, the tendency to rely on automation to perform the task is more pronounced when operators are removed from the control loop and asked to monitor its behavior than when they actively interact with the automation. Parasuraman, Molloy, Moulou, and Hilburn (1996) found that operators who control the behavior of the automation through active participation are better able to detect system malfunctions than those who simply monitor the system's behavior. In one such study, Wickens and Kessel (1979) asked operators to track system dynamics using a pursuit display with and without a concurrent failure detection task; the tracking task was designed such that operators were either actively controlling the dynamics or passively monitoring an autopilot which controlled the system dynamics. The results showed that operators performed the detection task faster when they remained active in the control loop than when they tracked the system passively, and this benefit remained even when the concurrent failure detection task was added.

In fact, the degree to which a system is automated plays a significant role in the operator's ability to resume control of the system should the automation fail (Endsley & Kiris, 1995). Rather than implement automation in the traditional sense in which tasks are allocated given the current technological capabilities, focus is now turning to consider how relegating tasks to automation will impact the operator's ability to maintain an awareness of the system's properties, which is critical to both system safety and failure recovery. Kaber, Onal, and Endsley (1998) found greater response times were required for automation monitors to perform a task adequately than active controllers after an automated system malfunction. Subjects performed a teleoperation manipulation task by controlling a robotic arm using a SpaceBall; the level of control was varied from complete human manual control to full computer control. Failures were simulated periodically, at which time subjects were required to manually control the task. The results showed that under "normal" conditions, more precise performance was achieved when the automation performed the task, but the benefit of lower levels of automation in maintaining human involvement in the process when system recovery after failure was required outweighed the cost.

3.2 Automation Undertrust

Rather than trusting a system too much, some operators may undertrust the performance of the automated system because (1) it has failed before (Muir & Moray, 1996) or (2) it has produced a high number of "alarm false alarms" (Parasuraman & Riley, 1997; Sorkin, 1988). Alarms which present the operator with an alert when a pre-defined threshold is exceeded may be mistrusted, and in the worst cases, ignored if the thresholds are too sensitive (Billings, 1997). For example, Casey (1993) tells the anecdotal story of a prison inmate who escaped because the guards were conditioned to ignore the alarms which were constantly being set off by birds flying overhead. Parasuraman and Riley (1997) report that automated alerting systems in aircraft, such as the ground proximity warning system, which indicate to the pilot when the aircraft is too close to the ground, are sometimes disabled because of their propensity for false alarms. Unfortunately, the danger of missing a potentially catastrophic situation often requires that such a conservative limit be inherent in such systems (Billings, 1997).

The occurrence of undertrust in an automated system may be influenced by an overconfidence in one's own unaided judgement, e.g., when the operator assesses the accuracy of his performance as higher than is appropriate. Riley (1996) reports that if the operator assumes that s/he can perform the task better than the automation, s/he may not rely on it even at the times when it performs the task better than the operator. For instance, Entin (1998) found that subjects were overconfident in their ability to detect targets in highly degraded images of a real-world scene and undertrusted information provided by automated guidance symbology which marked what it recognized to be targets with red squares and presented supporting information for the decision. The accuracy of the automated cueing system was set to be high with a 90% hit rate and 10% false alarm rate, or low with a 90% hit rate and 40% false alarm rate. As would be expected, the results showed improved performance when automated advice was available for determining whether an object was a target or not, and performance with a high accuracy system was better than performance with the low accuracy system. More importantly, the data revealed that subjects were more likely to rely on their own judgements and less likely to rely on the advice provided by the automated system in making non-target decisions, even though the accuracy of the system in recognizing non-targets was above chance.

Liu, Fuld, and Wickens (1995) found a combination of overconfidence and automation undertrust when they asked subjects to assign "customers" to the shortest of three supermarket service lines actively (manually) or to monitor the scheduling performed by an automated system passively. Unbeknownst to the subject however was that the "automated system" was actually a recording of his own manual performance. The results showed a higher sensitivity for detecting errors when the subject controlled the system actively rather than passively, but more interestingly, subjects were more likely to trust their ability to monitor the system over the performance of the automation when the two were actually the same. A study by Kantowitz, Hanowski, and Kantowitz (1997) examining the level of accuracy required of a route guidance system showed that drivers were more likely to undertrust the automated advice in familiar settings in which they were more confident in their ability to successfully navigate.

3.3 Summary

The presentation of automated guidance information induces biases in the calibration of trust in the form of overtrust as exhibited by operator overreliance on the automated advice or undertrust in the form of an overconfidence in one's own abilities. The consequences of the former are errors in missing potentially important events [e.g., mistakes in the automation as reported by McFadden et al. (1998) and Mosier & Skitka (1996)] as well as a failure to adequately consider all the information available before selecting a course of action (Mosier et al., 1992, 1998; Wiener, 1980). On the other hand, a consequence of the latter is potentially a failure to respond to a legitimate warning only because the sensor has failed before (e.g., if it is too sensitive and alerts the operator to numerous false alarms (Casey, 1993; Sorkin, 1988).

4. ATTENTION AND TRUST BIASES TOGETHER

The presentation of highly accurate guidance information compels attention by simplifying the visual search space, filtering the abundance of information with which operators are besieged, and implicitly signaling the relative importance of the various display elements. When trust in the information provided by one source and the belief that that information is reliable guides attention away from another information source that is just as or more reliable than the one the operator attends to, biases in both attention and trust are manifest. This miscalibration may be a consequence of emphasizing certain display features (e.g., with the presentation of a highly accurate cue) such that operators *overtrust* the information provided by the system which allows performance of the task with less cognitive effort, and fail to distribute attention as thoroughly throughout the rest of the environment. Consequently, one source of information is overprocessed at the cost of another.

Instances in which attention is allocated in a way that does not reflect the true information value, or trust, in a source is provided by examining how operators respond to an invalid cue, i.e., one which guides operator's attention to an area which does not contain the target. In the basic attentional literature, Posner and Snyder (1974) modeled this behavior using data collected from subjects who were presented with cues that were reliable 80% of the time and cues that were reliable only 20% of the time. The results showed that when a cue was highly reliable (e.g., 80% of the time), a benefit was present for detecting the target, but a cost was incurred when the cue was incorrect. On the other hand, when the cue had a low reliability (e.g., correct 20% of the time), performance benefited when the cue was correct but more importantly, *no cost* was found when the cue was incorrect (Posner & Snyder, 1974).

Jonides (1981) examined these shifts in attention using a peripheral cue or a central cue. A cue (an arrowhead) indicated the location of a target (letters) appearing at one of eight potential locations evenly spaced around the circumference of a circle. The cue could appear in the periphery of the display 0.7º away from where the target would appear or in the center of the imaginary circle; the validity of the cue was only 12.5%. Subjects were instructed to pay attention to the cue (in which case, they were told that the validity of the cue was fairly low) or told to ignore the cue. The results showed that when subjects were told to attend to the cue, performance improved when the cue was correct, but it did not matter whether the cue was presented in the periphery of their field of view or in the center. However, when subjects were told to ignore the cue, an interaction between cue validity and cue type was present such that detection of the target when it was cued in the periphery improved when the cue was correct, but no such difference was found when targets were cued in the center of the display. In other words, the presentation of the peripheral cues was so compelling that subjects were unable to ignore it. Further experiments revealed that peripheral cues could capture attention even when subjects do not expect a cue in their periphery.

The attentional behavior observed in these studies provide examples of biases in both trust and attention. The subject exhibits a tendency to overrely on the hypothesis suggested by a cue, and as a consequence, attention is less focused on that region in space which is highlighted by the cue and more focused on the information presented by the cue. For example, Conejo and Wickens (1997), using an air-to-ground attack simulation, found that trust in information which cued pilots of target location in an air-to-ground targeting task was so great that it resulted in nontargets at that location being classified as targets. That is, despite the presence of trials when automation cued an incorrect target (or no target), the pilot still allocated less visual attention to analyzing the location of other objects in the environment when the true target was away from the inaccurate cue or to analyze the presence of objects in the environment in order to determine that no target was present.

In the presentation of guidance symbology, the more information the augmented reality imagery provides, the more compelling it will be and the more likely it is that the operator will pay closer attention to the symbology and reduce attention allocated to the rest of the visual scene. Gempler and Wickens (1998) found that when flight path predictors were presented to signal traffic avoidance information, the predictive guidance symbology was so compelling that pilots stopped attending to what the aircraft was currently doing and instead focused on what the aircraft was (unreliably) predicted to do in the future. On those few trials when the predictor symbol was in error, this attention focusing strategy got the pilots into trouble…losing separation with the aircraft whose flight path had been incorrectly predicted. This reliance upon the automated attention guidance and the magnitude of this “garden path” effect was greatest when the unaided task was most difficult.

However, a catastrophic loss in trust, i.e., direct evidence that automation is unreliable, serves to temporarily increase an operator’s attentional breadth. Using a paradigm similar to that used by Yeh et al. (1999), Merlo, Wickens, and Yeh (1999) simulated an automation failure by presenting an erroneous cue that guided attention to a location 180° away from the target’s actual position. Such an obvious failure of the automated system destroyed operators’ trust in the automated system to such a degree that subjects became less reliant on the cueing information on subsequent trials and less drawn to the information presented by the cue, thus preventing attentional tunneling. In fact, when the unexpected high priority event was presented with a cued object on the next trial, *all* the subjects detected the unexpected event rather than the cued event as was typical behavior exhibited in earlier trials, i.e., when the automation was 100% correct (see also Yeh et al., 1999). It should be noted that after a few trials, trust was regained in the cueing information, consistent with the theory set forth by Lee and Moray (1994).

Display design may mediate the joint effects of trust and attention biases. In fact, evidence suggests that these biases are caused by an increased level of realism depicted by the display, which not only increases the level of trust one attributes to the information provided by a simulation but also *implicitly* directs attention away from what may be more important in the display. This review will turn examine how display enhancements to increase the scene realism with which one explores the environment potentially influence calibrations of trust and attention and more importantly, the extent to which such calibrations are optimal (Theunissen, 1998). Increased computer technology has produced a variety of techniques to enhance the perceived realism in an attempt to improve interaction within simulators and virtual environments, e.g., the use of three-dimensional perspective and/or stereoscopic viewing versus two-dimensional viewing, multimodal interaction in which proprioceptive and kinesthetic feedback is presented to the user (Wickens & Baker, 1995). The two techniques we focus on here are:

(1) Increase the level of detail with which information is presented, e.g., increasing the amount of texture on the surface or drawing the environment with more polygons allowing for finer detail.

(2) Allow for interactivity with the system, i.e., active control in which the operator moves himself through the space versus passive control, in which the operator is moved to a designated point by an automated system. Motion may be one of the most important cues for enhancing realism (Rinalducci, 1996), and the inclusion of active control (i.e., self-motion) enhances awareness of the world and the location of oneself within the world (Gibson, 1986).

4.1 Increasing Scene Realism

The design of virtual environments and augmented reality displays prioritizes the conveyance of visual information over the other sensory modalities. In fact, much focus has centered around implementing as many visual features as possible in the simulated space in order to conform to visual features present in the real world in the hopes that increasing the similarity of the simulated environment to the real world environment will benefit learning of a task and improve interaction within the simulation (Caro, 1988; Dwyer, 1967). Cues which benefit the visual fidelity of the environment include adding color to a display or increasing the number of pictorial cues for depth (e.g., texture, relative size, and intensity). In terms of increasing the level of reality computationally, such techniques may require increasing the number of polygons with which a scene is generated or changing the number of databases from which to generate the terrain display. Unfortunately, the computational power required to increase the level of fidelity within a scene constrains how rapidly an image can be rendered and may result in discontinuous movement and optical distortions, both of which lead to poor perception of the size of the environment (Henry & Furness, 1993; Péruch & Gaunet, 1998).

Increasing the scene detail in virtual reality environments is motivated by the attempt to improve interaction to such a point that the user acts as if he is interacting with the real world; that is, the more realistic the simulation, the more realistic will be the resulting interaction (Hendrix & Barfield, 1995). Realism can be enhanced by the use of head-tracking (Hendrix & Barfield, 1995) or stereoscopic viewing (Padmos & Milders, 1992) but it is not clear what role pictorial realism plays in creating a sense of "reality" (Wickens & Baker, 1995) or in teaching geographical knowledge about an environment. In this section, we briefly examine the effects of increasing scene detail on performance before discussing more extensively how incorporating this feature of realism may induce biases in attention and trust.

4.1.1 Performance Benefits for Increasing Scene Detail

Table 4.1.1.1 summarizes the studies which have examined the influences of increasing scene detail on performance and learning. In the table, the different techniques which have been used to enhance scene detail are specified. A "+" sign indicates a positive influence, "0" indicates no difference, and "-" represents a negative influence.

In first examining the performance effects of increased scene detail as shown in Table 4.1.1, the results of the literature show greater sensitivity for detecting depth changes (Barfield, Rosenberg, & Kraft, 1990; Kleiss, Curry & Hubbard, 1988) and locating targets (Bossi, Ward, Parkes, & Howarth, 1997; MacMillan, Entin, & Serfaty, 1994) when the visual scene is enhanced. In identifying depth changes, Kleiss (1995) found that what was important was the degree to which scene detail enhances the presence of absence of vertical information in the scene or the presentation of object size or density. In the case of visual search, the mediating factor was subjects' reliance on high quality visual information.

Less obvious are the effects of scene detail on the precision and control of one's navigation through the simulated environment; in this case, the findings show that how scene detail is increased within a display influences performance differently. Benefits are found when scene detail is enhanced by increasing the texture in the scene (Jaeger, 1998), generating the scene with a greater number of polygons (Johnson & Schroeder, 1995), and increasing the resolution with which the scene is depicted (Mann, 1987), with mixed results for increasing the scene density by increasing the number of objects present (Lintern & Garrison, 1992; Lintern & Koonce, 1991, 1992; Lintern, Sheppard, Parker, Yates, & Nolan, 1989; Lintern, Taylor, Koonce, Kaiser, & Morrison, 1997; Lintern, Thomley-Yates, Nelson, & Roscoe, 1987).

Table 4.1.1.1. Biases induced and performance effects of increasing the level of detail (+ = benefit; 0 = no difference; - = cost).

	Performance			Training	
	Distance Perception	**Precision of Control**	**Visual Search**	**Skill Acquisition**	**Knowledge Acquisition**
Increased scene detail					+ Wilson, Foreman, & Tlauka (1997)
Increased perspective cues	0 Surdick, Davis, King, & Hodges (1997)				
• Number of objects		+ Lintern et al. (1987) + Lintern et al. (1989) 0 Lintern & Koonce (1992) - Lintern & Garrison (1992) - Lintern et al. (1997)		+ Lintern & Koonce (1992) + Lintern et al. (1987) 0 Lintern & Garrison (1992) 0 Devlin & Bernstein (1997) - Lintern et al. (1997)	
• Increased texture	+ Barfield, Rosenberg, & Kraft (1990) + Kleiss, Curry, & Hubbard (1988)	+ Jaeger (1998)			+ Jaeger (1998) - Dwyer (1967)
Greater number of polygons		+ Johnson & Schroeder (1995)			
Increased resolution		+ Bossi et al. (1997) + Mann (1987)	+ Bossi et al. (1997) + MacMIllan, Entin, & Serfaty (1994)		
Dimensionality (3D vs. 2D)					0 Williams, Hutchinson, & Wickens (1996)

One can not examine the performance benefits of scene detail without noticing the abundance of research in the training domain which explores the required visual fidelity for the development of simulators used to teach a specific skill or mission. In this context, the results of performance during training of the skills using different levels of scene detail are task dependent: high levels of scene detail were beneficial for teaching air-to-ground attacks for bombing runs (Lintern et al., 1987, 1989) but did not significantly improve training for landing tasks, suggesting that the necessary visual cues relevant to the task were already present in the low detailed scene (Lintern & Garrison, 1992; Lintern et al., 1997). The results for transfer performance showed a similar trend. Training with a low detailed scene was more beneficial than training with a high detailed scene when performing actual landings, suggesting that the cues which pilots used when learning the skill in the moderate detailed scene were not present during transfer. In the case of mission rehearsal to aid environmental knowledge acquisition, the findings here, while mixed, suggest that the success of increasing scene fidelity is the degree to which the added detail (e.g., texture) makes certain aspects of the environment distinct (Wilson et al., 1997). In fact, the performance benefits as scene detail is increased may be described by an inverted U-shaped function such that adding detail initially may benefit performance, especially in aiding wayfinding relative to a scene with no visual cues (Jaeger, 1998), but at some point, the maximum benefits are reaped and the addition of more scene detail will not improve performance any more (Surdick et al., 1997; Williams et al., 1996), and may even impair performance by drawing attention away from the more critical details in the scene or diagram (Dwyer, 1967). Thus, the benefits of display reality enhancements may become apparent when the task relevant scene properties are taken into consideration and dependent on the extent that they are perceived. That is, simply increasing scene detail within a visual simulation may not guarantee optimal or improved performance (Padmos & Milders, 1992; Rinalducci, 1996).

4.1.2 Visual Realism and Attention Allocation

There is growing concern that increasing the level of realism may result in unintended consequences ranging from unintended use to inappropriate allocation of attention and trust in the computer generated imagery. Theunissen (1998) has noted that the presentation of a realistically simulated out-the-window view on an aircraft display to prevent controlled flight into terrain may lead pilots to attend to the display not only for navigational information as intended but also as a source of guidance information. Given that the data in the terrain database was not designed for guidance purposes, it is impractical to assume that the information will be error free, and while an examination of the accuracy of terrain databases so far has found less than a 1% error rate in the terrain databases, only 25% of the databases have been checked (Hughes, 1998). More importantly, a small error – no matter how rare – could lead to disastrous consequences when that information is used inappropriately for guidance.

Theories of information processing have examined how increased scene detail might affect attention allocation and note that increasing scene detail may result in the presentation of visual information which is distracting and which may potentially interfere with learning rather than enhance it (Attneave, 1964). That is, the presentation of highly realistic visual cues, which may be unnecessary for learning, induce an attentional bias. Support for this hypothesis was found by Dwyer (1967), when he examined whether the use of increased level of detail in illustrations influence learning. Students listened to a 40-minute oral presentation accompanied by a sequence of thirty-nine black and white slides; the degree of realism (detail) in the sequence of illustrations was varied between groups such that students viewed no illustrations, abstract linear representations, more detailed shaded drawings, or realistic photographs. Students were then asked to take four post tests consisting of identifying parts highlighted on a model, terminology, drawing, and comprehension. In general, the results showed that the realistic photograph was not as effective as an abstract linear representation. In fact, as the number of information cues increased, learning may have been reduced due to (1) a filtering in the central nervous system which prevented all the information in the realistic stimuli from being processed (Broadbent, 1965), or (2) simply perhaps the critical details were missing from the high resolution displays.

Thus, limitations inherent in the human visual system may make it impossible for the user to process all the visual cues present in the display. In listing several constraints of the visual system that can be used to guide the design of graphics displays, Gossweiler (1994) highlights the roll of attention. Given that tunnel vision results when the user focuses on specific objects, items which are not being attended to can be presented in less detail. Unfortunately, developing displays which capitalize on the human's perceptual limitations and enhance the visual quality in one location but not in another may implicitly guide the focus of attention towards only those areas which are presented in a high degree of detail. That is, visually enhancing certain regions of a display may guide attention to information in that location and away from critical information presented outside the region. Bossi, Ward, Parkes, and Howarth (1997) examined the use of a visual enhancement system, which illuminated information in the center of a driving simulator display, on driving behavior. Subjects were shown videotaped footages of a driving scenes, recorded in two luminance levels (high and low) and asked to track the vehicle located in the center while detecting the presence of artificial targets (landolt Cs) presented in one of four orientations (gap up, down, right, left) in the periphery. The results showed that the use of a vision enhancement system caused poorer performance in the detection and identification of artificial targets when the scene was presented at a low luminance level (e.g., driving at night) but not when the scene was presented at high illuminations (e.g., driving at dusk). That is, when the contrast difference between the center of the display and the information in the periphery was high, the richness of the visually enhanced information in the center of the display was more compelling, aiding performance on the tracking task, but such a benefit was obtained at the cost of detecting targets presented in the periphery.

Thus, the results of Bossi et al. (1997) provide evidence that attention allocation calibration may be influenced by the image quality with which the scene is depicted and strongly suggests that attention is drawn to high quality visual information. Such attention allocation occurs at the expense of information which may be presented at a lower quality but may be just as critical to the operator's task – in the case of Bossi et al., this information consisted of the presentation of targets in the periphery. The concern then is how trust is modulated as the realism of a depiction increases.

4.1.3 Trust Biases Induced by Scene Detail

It is likely that as scene fidelity increases, operators are more likely to trust the information provided, *even at times when such trust is undesirable* (Grieg et al., 1997; Salas et al., 1998; Suter & Nüesch, 1995; Theunissen, 1998). This is evidenced by Theunissen's hypothesis that pilots would use a simulated display as their primary flight display due to its high visual fidelity rather than as a secondary display complementary to the real-world view. In this case, the use of highly realistic visual imagery could compel the operator's attention away from information that is more critical to the operator's task, e.g., the out-the-window view, to computer-generated imagery which has the potential for displaying incorrect data (Hughes, 1998).

In an investigation of how scene detail influenced subjective ratings in an architectural scenario, Kaplan, Kaplan, and Deardorff (1974) showed subjects low- or high-detailed models of housing developments and asked subjects to rate the two in terms of functionality characteristics (e.g., privacy). Subjective ratings showed that subjects who were architecture students, i.e., domain experts, preferred the high detail model to the low detailed one even though there were no differences between the high and low detail models in the judgments made. Additionally, when subjects viewed the real architectural structures represented by the models, they reported that what they had seen was a satisfactory representation. While the researchers did not explicitly measure "trust" in accuracy, we may infer greater trust from students' higher preference for the high detailed model.

The issue of trust in visual imagery has been examined more extensively in the context of battlefield management by examining operator interaction with sophisticated automated target recognition systems which attempt to locate objects in cluttered scenes. The imagery with which these scenes are presented to the operator is often very degraded; hence, the system can help the human perform the difficult task of visual search. However, these systems are not yet capable of performing the difficult task of pattern recognition and therefore serve as an aid to the human operator to <u>detect</u> possible enemy locations and the presence of objects such as tanks and artillery which may be concealed by terrain or vegetation or covered by buildings. The system highlights the locations where it believes a target may be present and guides the operator's attention to that location. At that point, the operator must inspect the highlighted location and determine whether a target is present based on his own knowledge as to what the target looks like and the estimate of confidence that the automated system gives that it believes it has indeed located a target.

Ideally, when automated target recognition is provided, the operator performs the target detection task by extracting the critical information from the numeric data provided by automation (i.e., the confidence level with which the automation identifies an object as a target) and combining it with the visual information contained in the image itself at the cued location. Operators are successful at this task when the image quality is high, but as the image quality is degraded, more weight may be given to the visual information rather than the supplementary guidance information, i.e., a bias towards undertrusting the presentation of reliability (confidence) of the automation's inferences at that location. MacMillan, Entin, and Serfaty (1994) explored this hypothesis by assessing operator's trust of information provided by automatic target recognition systems with respect to variations in image quality. Subjects were asked to perform a range of tasks which included detecting the presence of targets (black and white silhouettes of tanks, trucks, and armored personnel carriers) against a blank background, discriminating targets from non-targets (silhouettes of trees, rocks, and buildings), detecting the presence of multiple targets contained in a visual scene, and detecting and discriminating targets from non-targets when supplementary numeric information to simulate a judgement value was provided by the automated system. The image quality of the targets was varied at three different levels. Not surprisingly, decreasing the image quality decreased subjects' ability to detect the targets, but the rate at which performance degraded with small degradations of image quality was more rapid than expected. When supplementary information regarding the automation's confidence in its identification of an object as a target or non-target was provided, subjects showed an overreliance on the visual image of the object, rather than the automation guidance provided by the supplementary numeric information, especially when the image quality was poor.

The pattern of results is consistent with behavior predicted in the trust literature when operators *undertrust* the information provided by an automated system (Lee & Moray, 1994; Parasuraman & Riley, 1997) and provide evidence suggesting that the calibration of trust in the information provided by automation is influenced by image quality. Operators rely more on visual information presented in a highly realistic scene than one with less reliability and consequently undervalue automated guidance information when the image quality is poor, even when they know that the accuracy of the image generation system is high. Additionally, operators may be overconfident in their ability to detect targets, even when automation is provided (e.g., a target recognition system as used by MacMillan et al., 1994), and this problem is exacerbated when they *undertrust* the automated data because the visual imagery is degraded.

4.2 Interactivity

The level of interactivity, or control, e.g., active vs. passive viewing, that a user is given in a system has been cited as a critical factor that influence the user's perception of the simulation's realism (Sheridan, 1996). In the design of virtual environments, the incorporation of active control of one's movements has been important in contributing to a sense of presence and realism within the simulation (Sheridan, 1996; Wickens & Baker, 1995), even playing a greater role than increasing scene detail (Welch et al., 1996). Active control increases the display's reality by keeping the user in the loop, constantly interacting with the simulation. Such benefits for active viewing may be in drawing attention to different areas of the visual scene, so that the user is in control of the "world". Consider, for example, the difference in knowledge gained about an unfamiliar city when one is the driver as compared to being the passenger. In fact, motion may be one of the most important cues for enhancing realism (Rinalducci, 1996), and the inclusion of active control (i.e., self-motion) enhances awareness of the world and the location of oneself within the world (Gibson, 1986).

Little research has been conducted to determine the effects of interactivity in terms of increasing trust in the system. Instead, the literature we will review focuses on how interactivity, which is a feature that is implemented in new virtual reality applications, influences attention allocation as a consequence of its increased tasks demands. It is intuitive to assume that navigating actively through a space requires the user attend to guidance information in order to maintain control. Consequently, this increased workload resulting from interactivity could influence reliance on automated guidance systems and the amount of attention allocated to certain navigation-relevant objects within the environment.

The results of the literature comparing the benefits of active to passive perception on varying tasks are mixed: some find benefits for active participation (Foreman et al., 1990; Gale et al., 1990; Gibson, 1962; Held & Hein, 1963, Larish & Andersen, 1995; Péruch et al., 1995, Stappers, 1989; von Wright, 1957; Williams et al., 1996), others benefits for passive viewing (Arthur, Hancock, & Chrysler, 1997; Satalich, 1995; Williams et al., 1996), while a number of other studies find no advantage for one or the other (Ito & Matsunaga, 1990; Schwartz, Perey, & Azulay, 1975; Walk, Shepherd, & Miller, 1988). The following sections will first describe these results and discuss the advantages and disadvantages for each level of control in terms of how these results may be attributable to the unequal distribution of attention throughout the environment as a function of interactivity.

4.2.1 Benefits for Active Viewing

The ability to navigate through a virtual space allows one to apply a spatial metaphor to the contents of the virtual environment, providing for a naturalness in interaction such that users, proficient at moving through real space, can apply this knowledge to moving through virtual space to find the information needed (Vincow, 1998). The mental models users develop from systems determine their behavior within the simulated environment and allow them to behave appropriately or inappropriately (Carroll & Olson, 1988).

Comparisons in real space of representations developed from studying a map have shown that when subjects view a scene passively, e.g., when subjects are shown one static view of a virtual environment or a single exocentric viewpoint of the space, information may be stored in orientation specific ways favoring performance on tasks whose presentation of information matches the orientation in which information was learned (Arthur, Hendrix, & Chrysler, 1997). Such *alignment effects* can be eliminated by navigating through the real world as the multiple viewpoints one sees of the environment allow one to develop representations of the space from different perspectives, thereby facilitating use of that information in a variety of orientations (Evans & Pedzek, 1980; Levine, Jankovic, & Palij, 1982; Thorndyke & Hayes-Roth, 1982). Hancock, Hendrix, and Arthur (1997) found differences in the formation of a spatial representation of a virtual checkerboard environment containing four differently colored spheres when subjects learned the space through active navigation rather than passive viewing of the space through a fixed viewpoint in the center. Note that while the subjects in the passive group could not move through the space, they could turn their heads to obtain different views. The subjects were then presented with static views of the environment in twelve different orientations (from 0° to 330° in 30° increments) and asked to examine the configurations and judge whether it was possible or not. The results were such that the response times from subjects who viewed the space passively could be fit to a quadratic function, with a maximum at 180° while subjects who were allowed to navigate freely through the space did not show this increase in response latency.

While the benefits for active viewing reported so far may be largely attributable to differences in the information acquired from active traversal of space versus passive viewing of static scenes, the benefits for active control are robust and are reported even in conditions when passive subjects are allowed to explore the environment. In the tactile domain, Gibson (1962) reported that active touch, in which the observer explored an object’s surface, led to better discriminability of an object’s shape and better identification of the object, than passive touch, i.e., when the object was moved passively under the observer’s hands. Gibson’s experiment was replicated in the visual domain by Stappers (1989), who showed subjects occluded two-dimensional shapes and asked them to identify the shape by actively moving towards the object (with mouse input) or passively watching the stimulus image move back and forth. The results revealed that shape identification accuracy was higher when subjects actively controlled the rate of accretion or deletion; that is, active control aided form perception of occluded objects better than passive control.

Péruch, Vercher, and Gauthier (1995) found benefits for learning through active exploration when they examined the effects of navigational control for wayfinding. Subjects were immersed within a space containing four cubes of different colors, hidden behind walls. To learn the environment, subjects in one condition navigated freely through the environment using a joystick, others watched the path of another subjects navigating through the environment, and a third group simply viewed static scenes of the space. After being exposed to the environment, subjects were positioned within the virtual environment and asked to find the shortest path to a target cube. The results showed that subjects who actively explored the environment found paths faster and more accurate than those in the two passive exploration groups, and that there was no difference in performance between the two passive exploration conditions (static or moving), suggesting that active motor behavior (i.e., navigating with a joystick) was necessary for perception.

Active control may produce not only gains in spatial knowledge but also increased detection of dynamic spatial orientation changes. Larish and Andersen (1995) asked subjects to “fly” through a flight simulation either actively or passively; in the active condition, subjects followed a pre-determined path by performing either a compensatory tracking task (maintaining a constant heading) or a pursuit task (follow a moving target). At one point, the flight took the subjects into a cloud bank (i.e., a time when no visual information was displayed) and upon existing the cloud bank, subjects were asked to detect the change in orientation. The results showed a greater sensitivity in detecting change after active observation.

The task of active navigation may be a difficult one, and thus, it is conceivable that passive exploration can allow viewers to better develop an accurate representation of the world than active explorers since the passive observers can allocate all their cognitive resources to the spatial acquisition task. Williams et al. (1996) provide evidence, finding that the benefit for active navigation in learning the spatial layout of small scale environments were eliminated as workload increased. In their experiment, student pilots were given maps of a region and taught a pre-specified flight path; some pilots were given the opportunity to familiarize themselves with the region by flying the path actively, others viewed it passively, while a third group did not fly the route at all but just studied the map. Subjects were then asked to fly an evaluation flight, and upon reaching the final checkpoint, the map was removed, and subjects were asked to fly back to the initial point. Afterwards, subjects were shown a map of the region and asked to draw the flight route, checkmarks, and landmarks. The results showed that subjects who navigated through the terrain actively formed a more accurate mental representation than those who watched passively. However, when the speed of travel increased, the amount of workload required to fly the plane increased so much that less attention was paid to the terrain being traveled and the benefit for active viewing disappeared. We now consider the benefits for passive exploration.

4.2.2 Benefits for Passive Viewing

Satalich (1995) found that constraining navigation by asking subjects to follow a pre-defined path drawn in the virtual space drew attention away from objects of interest in the visual scene so that passive explorers showed better performance on tests of spatial orientation than active controllers. Subjects explored a virtual building containing 39 rooms and 500 landmarks (paintings and random objects displayed in the hallways) presented using an opaque HMD. Subjects could navigate actively through the building freely or were asked to stay along a specified path; in the passive condition, subjects were moved at a constant speed along a path. Measures of spatial orientation (collected by asking subjects to point to objects which were not presently visible, estimate route and Euclidean distances between current position and an object or between two objects, and navigate between two points) showed higher performance for the passive guided group than the active guided group. Members who explored the scene actively were too concerned about following the designated path and failed to look at the configuration of objects in the environment.

While passive viewing of a scene may limit the information one acquires, at times this limited information is all that is necessary. Arthur, Hancock, and Chrysler (1997) attributed benefits found for the passive viewing of a single fixed viewpoint of a virtual room relative to active exploration to the nature of the task used to measure performance. In their investigation, subjects were randomly assigned to one of three viewing conditions, differing in the type of exploration available. Subjects could either navigate freely through a real room, navigate freely through a virtual room, or passively view the virtual room through a single fixed viewpoint presented monocularly. Each room contained nine common objects, placed on the floor. Subjects were given as much time as they liked to view the configuration of objects in the room and were then asked to complete a map of the space, on which two of the nine objects were present to provide subjects with the scale and orientation for the map. The results showed that subjects in the fixed viewing condition performed better on the map drawing task than subjects in the other two conditions, a result the authors attributed to the nature of the map drawing task, in which subjects in the fixed viewpoint condition viewed and drew the objects from that single orientation.

4.2.3 Mediating Factors: The Role of Attention Allocation

The conflicting results between experiments comparing active and passive navigation may be attributed to attentional differences regarding which objects one attends to, the type of information available in the environment and sensitivity to that information, or to differences in the ease with which changes in the form or shape of the object can be perceived (Flach, 1990; Ito & Matsunaga, 1990). In studies in which experimenters emphasize that subjects attend to specific features of the task or environment, no differences are found. For example, Shebilske et al. (1998) found no difference between active and passive control in learning flight skills when subjects trained in pairs with one member of the pair controlling the task and the second member of the pair observing as the nature of the social interaction guided the attention of both members of the pair to the critical task features. In tasks of navigation and wayfinding, Wilson, Foreman, Gillett, and Stanton (1997) found no difference between active and passive navigation when they emphasized to passive viewers the importance of attending to the spatial properties of the virtual environment.

The relevance of the context in which the objects of interest are presented to the task being performed may influence the nature of results when comparing active to passive interaction. Wallis and Bülthoff (1998) presented subjects with a virtual scene of a suburban road and allowed them to interact with the scene actively (i.e., subjects drove through the scene) or passively (i.e., subjects were driven through the scene or shown static images of the scene). Subjects were then shown the same scene and asked to detect changes in terms of object shape, location, presence or absence, or color. The results showed that overall, detecting changes in the visual scene was degraded when subjects moved through the scene, whether this movement occurred actively or passively. However, active driving benefited the detection of object changes along the roadway, i.e., for those objects relevant to the driving task, leading the authors to suggest that the results could simply be a consequence of task dependent influences on attention.

A final factor to consider is how the additional task of active exploration – and the workload it imposes – may influence where attention is allocated within a virtual environment. Evidence suggests that increased workload induces trust in automation (McFadden et al., 1998; Riley, 1994, 1996); consequently, when automated guidance information is provided, active explorers may exhibit a greater “complacency” effect when the automation fails and a greater attentional tunneling cost (an attentional bias) when it does not.

Thus, performance differences when examining active versus passive interaction may simply be the result of differences in the type of information one hopes to gain based on the task at hand. In other words, the very task of navigation calls attention to certain features of the environment which would otherwise be ignored, e.g., the driver of a car is more aware than the passenger of obstacles in his path, such as an approaching vehicle in the next lane, and costs or benefits to active navigation may simply be attributable to the fact that what is measured in terms of knowledge acquisition may not have been what was required by the active control task. If this were the case, then we would expect that if subjects who perform the active navigation task perform differently than subjects who view a scene passively, the results are attributable to the fact that the task requires them to focus attention on different areas of the visual scene. The current study seeks to provide evidence in support of this hypothesis.

5. RESEARCH QUESTION

The literature review has provided evidence for four display features which lead to departures from optimality in calibrating attention and trust: (1) superimposing symbology onto the far domain, (2) displaying a visual scene using an egocentric view, (3) providing intelligent cueing information to guide attention throughout the scene, and (4) varying the reliability of this cue. In the current study, we propose to manipulate two of these four variables. One variable that is expected to produce an attentional bias – i.e., cueing, and a second variable that will produce a trust bias – i.e., reliability. Additionally, we propose to examine two additional display features which could potentially influence these calibrations: enhancing the image quality with which the visual scene is depicted may induce an attention bias and increase trust in the information provided, and increasing interactivity which may produce attentional effects given the task-driven attention requirements.

The most prominent finding from the automation literature regarding operator behavior when guidance information is provided is a cost-benefit trade-off imposed by cueing, such that cued objects are assisted in their detection but direct attention away from other sources of information which either should be detected (Conejo & Wickens, 1997; Ockerman & Pritchett, 1998; Yeh, Wickens, & Seagull, 1998) or which may assist in diagnosing the situation (Mosier et al., 1998). Furthermore, such a cost may be amplified by the location or FOR of the cue as was shown by Yeh and Wickens (1998) in which head-up conformal symbology which represented the target’s location impaired detection of the unexpected target relative to the head-down non-conformal symbology. This difference may be attributed to the apparent realism of the cue. That is, the augmented reality cueing in the HMD induced a sort of attentional tunneling that was reduced when cueing was presented in its less “real” form on the hand-held display.

This tunneling may be mediated by the reliability of the cueing information, such that the presentation of partially reliable cueing symbology would expand one's attentional breadth (Merlo et al., 1999). Such an effect would be attributable to decreased trust in the cue and consequently, less attention allocated to that information source. However, the presentation of cueing information is compelling and may induce other biases in trust when the validity of the cue is less than 100% reliable, i.e., an increase in the rate of false alarms when using the existence of a cue to trigger a "target" response and ignoring the appearance of the "raw data" beneath. Conejo and Wickens (1997) found that operators trusted and followed the automation even when information in the environment contradicted the information provided by the automation-based cueing, and Gempler and Wickens (1998) found that predictive guidance information was so compelling that pilots paid more attention to what the predictor suggested the aircraft would do in the future – even when this information was unreliable – instead of focusing on where the aircraft was currently headed. In both these cases, attention appeared to be less focused on the region in space which was cued than on the information *suggested* by the cue, i.e., subjects relied on the automation and allowed it to perform the perceptual work of target identification, even when it did so incorrectly.

It could be argued that once the cue is found to be partially reliable, it will simply be ignored. However, evidence suggests that trust, *if* lost, is regained after a few trials of accurate cueing (e.g., Davison & Wickens, 1999; Merlo et al., 1999; Parasuraman & Riley, 1997; Riley, 1994, 1996). Furthermore, the results of Conejo and Wickens (1997) and Gempler and Wickens (1998) suggest that in some instances, when the unaided task is difficult (i.e., processing of the "raw data" is hard), reduced cue reliability may not influence subjects' attention allocation strategies since performance of the task with supplementary cueing information – as long as its reliability is relatively high – is believed to be better than performance achieved without the automation. This finding is consistent with the model of trust proposed by Lee and Moray (1994), in which operators were more likely to allocate tasks to an automated system because of a decrease in the operator's confidence in his own abilities.

Despite the abundance of research examining performance effects with greater levels of reality, few studies have empirically investigated how the distribution of attention within a display is affected by scene detail or active navigation, or how the reality with which the information is displayed biases operators to trust one data source over another. However, both these variables could plausibly induce attention effects and modulate trust effects (Theunissen, 1998). If, for instance, greater image reality does indeed induce greater trust in the simulation display – and inherently, the data used to generate that visualization, operators may be less likely to detect errors in the terrain data than when these errors are presented using a less compelling information source (e.g., a paper map). Additionally, as applications for augmented reality are being developed to superimpose symbology over what may potentially be a computer generated scene so that operators can perform tasks in hostile or hazardous environments remotely (Drascic & Milgram, 1996) or so that maintenance engineers can perform their quality inspection tasks more efficiently (Ockerman & Pritchett, 1998), the issue of what level of detail a scene needs to be generated and whether this detail influences the operator's reliance on the information provided by an automated system becomes an issue. For example, MacMillan et al. (1994) and Entin et al. (1998) show that image degradation influences trust and find that when the image quality is poor, operators *undertrust* the cueing information imposed to augment the degraded imagery; a logical extension of this finding is that high fidelity image quality would lead one to overtrust the imagery. Thus, a hypothesis to be set forth for investigation is that any guidance cue associated with a more realistic scene may be trusted more and receive more attention even when more important information is present in areas uncued by the guidance information.

Finally, we propose that increasing interactivity which increases the realism of the scene may indirectly induce attention biases as workload is also increased. Consequently, the operator, actively navigating through the scene will be allocating attention more on the guidance information, may become more dependent on the automated cueing information than the passive explorer, and may therefore allocate attention in a non-optimal fashion to what is cued, even if active navigation does not actually increase trust in the cue. It is believed that the task of active navigation will benefit the detection of those objects which are directly relevant to the task of navigating (Wallis & Bulthoff, 1998), and therefore, it is anticipated that the detection of target objects presented along the path will be greater for the active navigators than for the passive viewers.

In order to investigate these issues, subjects were asked to navigate through a simulated terrain environment searching for military targets camouflaged in the terrain in a scenario in which they were viewing a display from an unmanned ground vehicle. Terrain visualization was assumed in the simulation to be rendered from data recorded and presented by sensors. To aid the subject in the target detection task, automation attempted to interpret what objects (i.e., targets) existed in the environment and when possible, provided a cue to augment the terrain display to guide the user's attention to target locations. However, at times sensors were unreliable and guided the user's attention to an incorrect object (e.g., cueing an object which was not really a target). More importantly, there were objects in the environment that the sensors were unable to detect; these objects were of high priority and uncued, and detection of these targets would provide data as to the degree of an operator's reliance on the cueing information.

Within the context of this general paradigm, two variables were manipulated orthogonally: scene detail and interactivity. For the first of these variables, the level of fidelity with which the environment was presented was varied in terms of the degree of texture detail which defined the surfaces. For the second variable, navigation either occurred actively, in which the subject moved himself through the space, or passively, in which the subject was moved through the environment. We were interested in whether these variables would influence attention allocation and trust attribution in the visual scene and how these techniques for enhancing reality would modulate the effectiveness of the target cues.

6. METHOD

6.1 Subjects

Sixteen subjects (15 male, 1 female) participated in the experiment. 13 were members of the US Army or Army National Guard; 3 were members of the US Marine Corps. Subjects had an average of 9.9 years of military experience, ranging between 3 and 13 years. 3 served as captains, 1 as a midshipman, 11 were sergeants, and 1 was a specialist.

6.2 Task Overview

The task performed by subjects consisted of two stages: (a) target detection and (b) target azimuth reporting. The target *detection* task required subjects to scan the display looking for any one of four target objects (tanks, soldiers, land mines, and nuclear devices). In order to increase the difficulty of the target detection task, distractor objects (e.g., trucks and trees) were presented in conjunction with the target objects. The ratio of targets to distractors was 4:1, i.e., the likelihood that a given object was a target was 80%. Unbeknownst to the subjects, with the exception of the nuclear device, only one target could be visible at any given time, and the time during which it was visible was defined as a "trial".

Of the target trials, three of the objects (tank, soldier, and land mine) were presented 90% of the time (30% each) and were therefore expected; the fourth (nuclear device) was presented only 10% of the time, concurrently with either a cued or uncued target (soldier), and was unexpected. Subjects were not told which target to search for but were instructed that reporting the nuclear device took precedence over the detection of all other targets. Once the target was found, the subject was required to report the target's *heading*, the current compass direction of the target with respect to his current location. Subjects were instructed not to report the presence of distractor objects.

In this study, cueing was provided to the location of all tanks, half of the soldiers, and half the land mines. Two levels of cue reliability were varied between subjects: one which was 100% reliable and the second which was 75% reliable; subjects were informed as to the level of reliability of the cueing symbology. The manipulation of reliability was implemented through the cueing of distractors: when cueing was 100% reliable, distractor objects were never cued, but when the cueing was only 75% reliable, *every distractor object that was present was cued*. Comparison of the data from the two conditions were then used to determine whether trust in the system was higher than warranted given this system unreliability.

While performing the target detection task, subjects were also asked to perform a terrain association task to assess how realism influenced attention allocation as well as the degree of trust placed in the potentially unreliable terrain simulation. Subjects were asked to monitor their position on a paper map as they moved through the environment and respond to those instances in which the visual display presented information contradictory to the more accurate paper map.

6.3 Apparatus

The experiment was conducted using an immersive virtual reality system, the ImmersaDesk, a drafting table style 4’ x 5.5’ projection based display. Subjects sat 33” away from the ImmersaDesk at an eye height of 24” so that the display subtended an area of approximately 90º visual angle. Subjects were asked to wear shutter glasses, on which a head tracker was attached.

6.4 Displays/Tasks

The terrain was developed using geographical data of the National Training Center located at Fort Irwin, California. The level of realism with which the environment was presented was varied in terms of the number of polygons used to generate the terrain and the degree of texture detail which defined the surfaces. Figure 6.4.1a depicts a high detail scene; Figure 6.4.1b depicts a low detail scene.

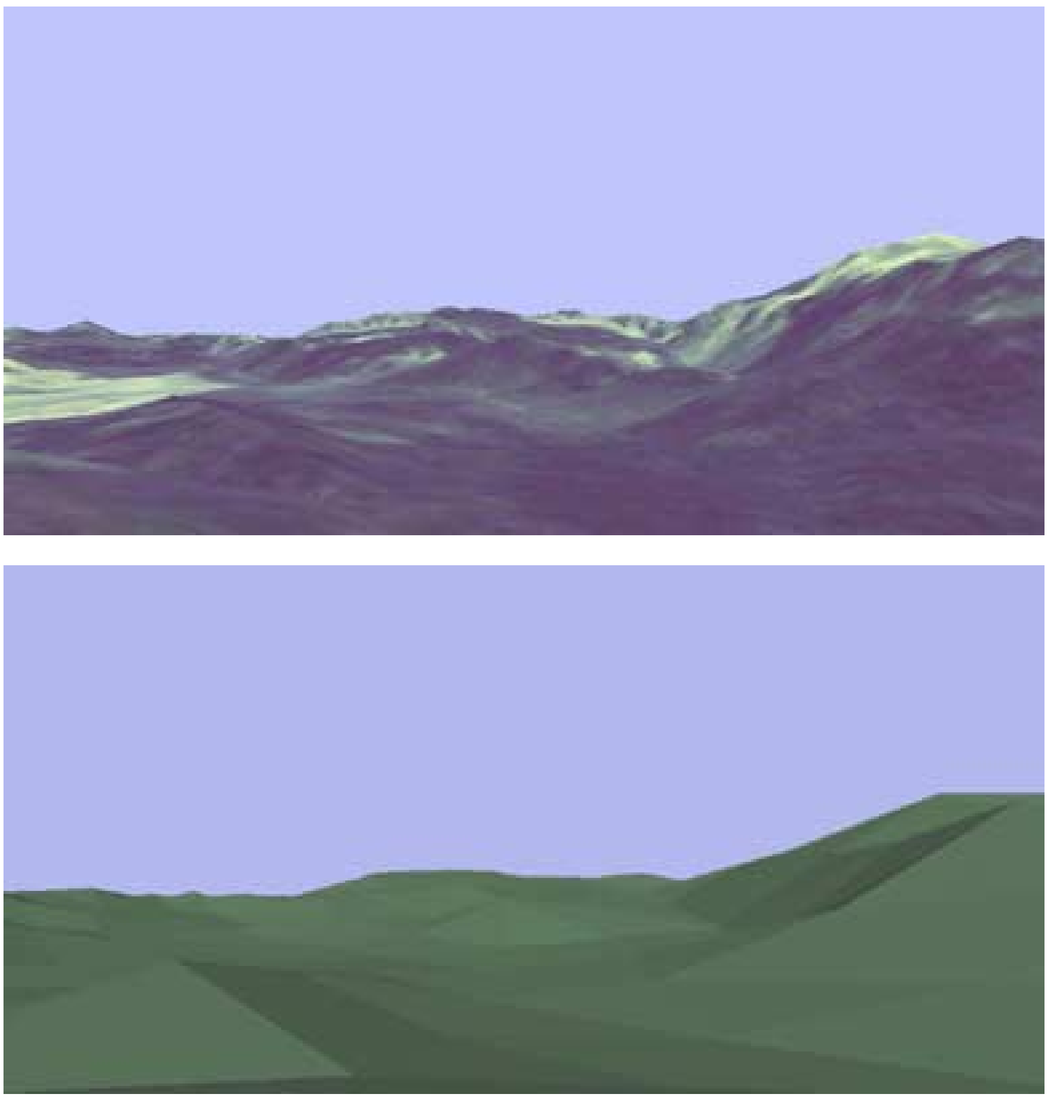


Figure 6.4.1. (a) high detail scene (b) low detail scene.

The number of polygons in the high detail terrain was approximately 5:1 relative to the number of polygons in the low detail terrain. The high detail scene, shown in Figure 6.4.1a, was rendered with approximately 20,000 polygons and textured with a satellite image of the terrain. A noise image was superimposed onto the satellite image which modulated the intensity of the underlying satellite data, hence adding more detail to the scene and enhancing the realism of the terrain. This detail texture was applied in such a way that finer texture indicated points further away.

The low detail scene, shown in Figure 6.4.1b, was created with 4,000 polygons and flat-shaded. That is, the polygons were shaded in green but the intensity of the color was dependent on the angle between the polygon's normal and a light source, located at a 45° angle above the horizon in the southern sky. The smaller the angle, the brighter the green the polygon would appear. Note that in the low detail terrain, each polygon was uniformly colored over its surface.

The distribution of polygons within each scene also differed. In the low detail terrain, the number of polygons with which the terrain was depicted was dependent on the distance from the user's position so that terrain close to the user was drawn with a greater number of polygons than terrain farther away. In the high detail terrain, however, the polygons were drawn in such a way that they more accurately conformed to the underlying data, e.g., mountains were generated with a greater number of polygons, and hence, greater detail, than flatter terrain.

The two scenes were presented at a resolution of 1024x768 and at a frame rate of 24Hz.

Stimuli, shown in Figure 6.4.2, were placed in the terrain.

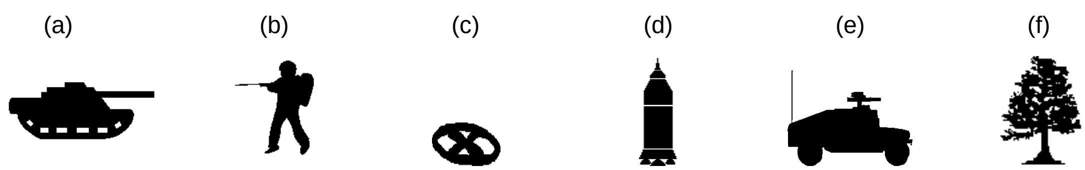


Figure 6.4.2. Stimuli. Target objects consisted of: (a) tank, (b) soldier, (c) land mine, and (d) nuclear device. Distractor objects were: (e) truck and (f) tree.

The tank, soldier, land mine, and nuclear device [Figure 6.4.2(a) – (d)] served as target objects; the truck and tree [Figure 6.4.2(e)-(f)] were placed in the environment as distractors. Prior research has revealed that the land mine, with its smaller visual angle, is less visible and salient, and hence will be considered the low salience target (Merlo et al., 1999; Yeh et al., 1999). Objects in the high detail terrain were camouflaged, i.e. colored in shades of brown, green, and black; in the low detail terrain, the targets were simply silhouetted images shaded green. Since the shading of the terrain varied, the intensity of the targets was adjusted adaptively at each location so that the contrast ratios between the target and the terrain were similar for all targets[*]. The greater salience of the nuclear device was insured by presenting them at a higher contrast ratio with the background than the other three targets. Additionally, the nuclear device was placed in the terrain at a location between the path and the soldier with which it was paired, i.e., at a smaller visual angle to the viewer. The location of tanks, half the soldiers, and half the land mines were cued with an arrow which signaled the current lateral and vertical location of a target.

The symbology is presented in Figure 6.4.3, which depicts the cueing reticle, heading tape, and tunnel pathway superimposed over the high detail scene.

[*] This adjustment was based on the performance and subjective assessment of 5 pilot subjects. Results of the pilot study are presented in Appendix 2.

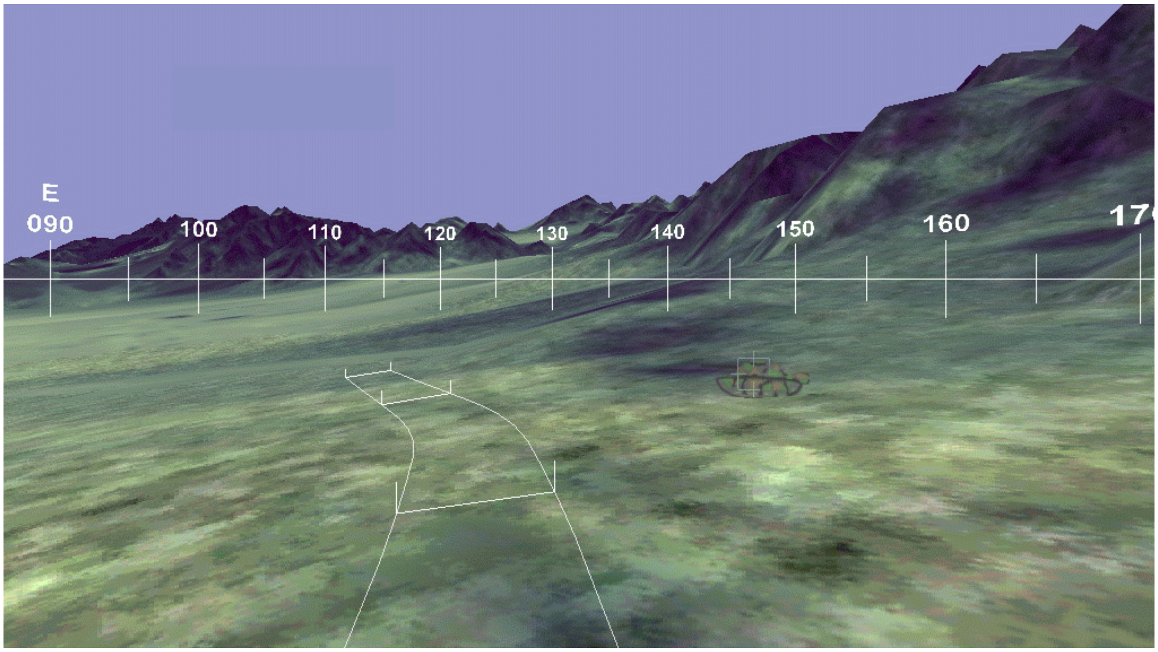


Figure 6.4.3. Cueing reticle, course deviation indicator, and heading tape superimposed in the high detail scene.

Symbology shown in Figure 6.4.3 was presented in white on the ImmersaDesk display.

When target objects were cued, a reticle consisting of a square with four crosshairs was presented conformally over the target to signal the current lateral and vertical location of a target with respect to the subject's head orientation. The reticle shown in Figure 6.4.3 is superimposed over a tank. For tanks and soldiers, the reticle appeared once the target was within 1500m of the subject's position; for land mines, the reticle appeared when the target was within 750m of the subject's position due to its smaller size relative to the tank and the soldier and hence, lower saliency at greater distances. The lock on reticle was of course *not* used to signal the presence of any uncued targets. The cueing symbology remained superimposed over the target until the target was detected or until the target was passed.

When cueing was only partially reliable, the cueing symbology sometimes indicated the location of a distractor object, i.e., either a truck or tree, but usually (75% of the time) signaled a correct target. Subjects were instructed not to detect the distractors, even if they were cued.

A heading scale to be used for azimuth judgment was presented in a decluttered format: when subjects detected the target by pressing the left or right wand button, the heading tape appeared, conformal to the horizon line. This case is depicted in Figure 6.4.3. The four cardinal directions were marked on each heading tape.

Half the subjects were asked to direct the movement of the unmanned ground vehicle and *actively* move through the environment. The tunnel symbology, as shown in Figure 6.4.3, consisted of three inverted wickets connected by lines along the bottom corners. Each wicket was 50m high and 100m wide, with 750m separating one wicket from the next. The subjects' task was to closely track the path as commanded by the symbology as closely as possible. To control the rotation of the vehicle, subjects rotated the wand in the direction they wanted to turn. The speed was held constant at 250mph. Note that the tunnel symbology was displayed whether subjects were actively navigating through the terrain or not.

Subjects were also asked to perform a local positioning task. Subjects were given a military map of the terrain, on which 4 checkpoints were marked, and were instructed to report whenever one of these checkpoints was passed (accomplished with a button press on the wand). To aid them in their task, a message cueing the subject to the approaching checkpoint was presented when the subject was approximately 1000m away; this distance was varied randomly to prevent subjects from responding to the checkpoint task primarily based on the presentation of the message.

The checkpoints were not displayed on the simulation, forcing subjects to reference the paper map. Subjects were told that the information on the map was accurate but warned that at times the information in the terrain simulation could contradict the information on the paper map; should this occur, they needed to report the inconsistency to the experimenter. Such a scenario occurred in the last experimental block in order to assess trust in the computer image. In this case, the map showed a checkpoint located next to a mountain when no mountain was present on the display.

6.5 Experiment Design

The experiment was a mixed design as shown in Table 6.5.1.

Table 6.5.1. Experimental design.

			Within Subjects Scene Realism	
			High Detail	Low Detail
Between Subjects	Active	100% Reliable Partially Reliable (75%)	**TARGET TYPE** CUED: Tank, Soldier, Mine | UNCUED: Soldier, Mine | Nuclear Device ← *Expected* → | *Unexpected*	**TARGET TYPE**
	Passive	100% Reliable Partially Reliable (75%)	**TARGET TYPE**	**TARGET TYPE**

As Table 6.5.1 shows, the manipulations of cue reliability (100% reliable vs. partially reliable (75%)) and interactivity (active versus passive viewing) occurred between subjects. The manipulation of scene detail (high versus low) and target type (cued versus uncued targets, high versus low expectancy) occurred within subjects.

The six different paths (1 practice and 5 experimental) through Fort Irwin terrain that were used in the experiment are presented in Figure 6.5.1.

Each path constituted one block of trials; a trial can be defined as the presence of an object, whether it be a target or distractor. The experiment consisted of 2 practice blocks and 11 experimental blocks. Each path was presented twice during the experiment: once in high detail and once in low detail. For each level of scene detail, subjects were presented with one practice block, consisting of ten targets, and five experimental blocks, each containing a set of twenty targets. Once the subject completed the 10th block, they were instructed that they would view one more block, and at one point in this 11th block, the terrain depicted in the simulation contradicted the terrain information presented in the map (e.g., a mountain presented on the military map was displayed as flat terrain), thus simulating a case in which an error occurred in the terrain database.

In the practice block, subjects viewed a total of 10 targets and 3 distractors. The targets consisted of 3 tanks, 3 soldiers, and 3 land mines, each, and 1 nuclear device; the 3 distractor objects consisted of 2 trucks and 1 tree. Each experimental block consisted of a total of 25 objects: 20 were targets (6 each of tanks, soldiers, and land mines and 2 nuclear devices) and 5 were distractors (3 trucks and 2 trees). The target objects were distributed throughout the paths so that they were at least 750m apart**.** The exception was the presentation of the nuclear device, which appeared concurrently with a soldier. As it was an “unexpected” target, the nuclear device was presented within 15° of a cued or uncued soldier.

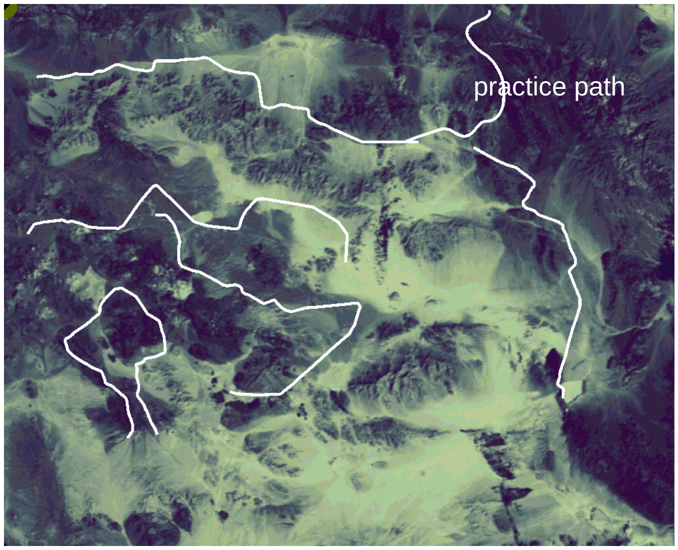


Figure 6.5.1. Paths used in the experiment.

Cueing was present 65% of the time: subjects were guided to the location of all tanks, half of the soldiers, and half of the land mines. When cueing was only 75% reliable, subjects were guided not only to the location of all the target objects but also to the location of all of the 5 distractor objects. To foster trust in the cueing information, subjects in the partial reliability condition were first presented with 3 experimental blocks in which the cueing information was 100% accurate, followed by 8 blocks in which the cueing information was only 75% accurate. In general, the data for these first 3 blocks were not included in the data analyses. However, the first inaccurate cue, encountered by subjects in the fourth block, will be significant for the signal detection analysis described below.

6.6 Procedure

The experiment took approximately 2.5 hours during which subjects were given the instructions for the experiment (see Appendix 3) and then performed the experiment. Subjects were first given a contrast sensitivity test; subjects scoring below 20/20 on this test were dismissed from the study.

Subjects were instructed to pretend that they were battlefield commanders viewing the display from the viewpoint of an unmanned ground vehicle; the display varied in its level of detail. Their primary task was to find the targets and send information back to their troop regarding the objects' position. They were told that automation was available to aid them in the target detection task and were informed as to the level of reliability of the automation. Their secondary task was to monitor their location on a paper map and call out checkpoints as they passed. Half the subjects actively navigated through the path, using heading information presented by tunnel pathway symbology. These paths were recorded and played back for the passive navigation subjects.

Subjects interacted with the display using a wand and shutter glasses. A diagram of the wand is presented in Figure 6.6.1.

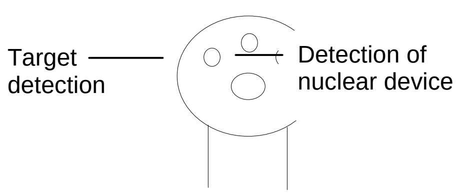


Figure 6.6.1. The wand.

The wand has three buttons and a pressure-sensitive joystick. Only the buttons were used during the experiment to make responses. The joystick was not used at all. Instead, the speed of travel through the environment was constrained to be 250mph. To move through the terrain, active navigators needed to rotate the wand in the direction in which they wish to move.

Subjects responded to the different tasks using the wand’s three buttons. To indicate that a tank, soldier, or land mine was detected, subjects pressed the left button on the wand. To indicate that a nuclear device was detected, subjects pressed the right button on the wand. Once the left or right button was selected, a heading scale appeared along the horizon line and subjects were asked to verbally report its location by calling out the target’s bearing.

Subjects responded to the local positioning secondary task by pressing the middle button on the wand whenever a checkpoint was passed.

Once subjects completed the five experimental blocks within one level of scene detail, they were asked to complete a questionnaire assessing three different aspects of the simulation:

(1) their perceived workload during the task (using the NASA-TLX scale),

(2) their perceived realism of the terrain information, and

(3) the amount of trust they placed in the cueing information and in the terrain information displayed.

At the end of the experiment, i.e., after subjects had adequately viewed both scenes, subjects were presented with a post-experiment questionnaire to directly assess the perceived realism in the two levels of scene fidelity. The questionnaires are presented in Appendix 4.

6.7 Performance Measures

The dependent variables collected from the primary target search task were the distance from the viewer to the target at the time of detection and the accuracy for target detection. In order to determine whether the symbology influenced the amount of scanning in the environment, data describing the amount of head movement along the x-, y-, and z- axes was collected. Additionally, data concerning the number of times and the amount of time the subject spent searching for information in the left, center, and right areas in the near and far distances of the display was collected. Note that the center points for the aforementioned view angles were at the center of the shutter glasses, rather than the center of the eyes.

The dependent variables collected from the secondary local positioning task were response time and accuracy for the detection of the checkpoints. Specifically, the duration of time that elapsed between the passing of a checkpoint and a button press on the wand was used. Implicit trust was also measured by asking subjects to report inconsistencies between the terrain database and the paper map, specifically the reporting of the missing mountain in the simulation.

For the active navigators, position tracking error was calculated using both root mean square error (RMSE) and mean absolute error (MAE) measures. The position of the remotely controlled unmanned ground vehicle was sampled every 0.5 seconds and errors were computed by comparing the current position with the target position. Lateral error was computed by comparing the vehicle’s current horizontal and longitudinal location with the target horizontal and longitudinal location.

Finally, subjective measurements as to the degree of realism of the terrain simulation and the amount of trust placed in the automated cueing information was collected.

7. RESULTS

Overview

The data were examined to determine the effects of cueing on attention allocation, the effects of cue reliability on trust, and how these effects were mediated by target expectancy, scene detail, and interactivity. Repeated measures analyses of variance (ANOVAs) were used to analyze much of the data in the experiment. The graphs are plotted with averages and error bars representing ±1 standard errors from the mean. Note that since subjects in the 75% reliable condition were presented with 100% reliable cueing for the first three blocks, the data from this group used for the ANOVAs were based only on the last seven blocks, when the cue was only 75% reliable.

Because we do not hypothesize that all the dependent variables would plausibly be influenced by all the independent variables (or if they were, such influences would not be of theoretical or practical interest), full ANOVAs on all dependent variables are not described in the text (these are included in Appendix 5). Instead, the data analyses focuses on the following:

7.1 Basic target detection findings, examining how cueing information, both perfectly and imperfectly reliable, influences attention allocation and how this behavior may be mediated by target expectancy and saliency.

7.2 Effects of image realism on target detection performance in order to determine how simulation fidelity guides attention allocation.

7.3 Effects of interactivity on target detection performance, exploring the issue as to whether performance differences in active and passive viewing, if they are present, can be attributable to where one focuses attention in the environment, and how this behavior may be influenced by the presentation of cueing information.

7.4 Effects of cue reliability on target detection performance (implied trust) to determine first, whether cueing information fosters trust (i.e., will subjects overtrust the automation and indicate that an object is a target just by the presence of a cue which overlays the object); secondly, how this trust bias influences attention allocation as measured by the detection of expected and unexpected objects; and finally, how these biases in attention and trust are influenced by interactivity and image realism.

7.5 Trust effects due to image realism: checkpoint detection task. This measure of implicit trust as calibrated through performance on the local positioning task was used to determine whether the compellingness of a high fidelity image would influence the trust one placed in that image.

7.6 Subjective ratings: workload and assessments of image reality

In each of these sections, the effects of the independent variables and how they are influenced by factors of target detection (i.e., target saliency, expectancy, and cueing) and implicit trust will be discussed.

7.1 Target Detection

The benefits of cueing on target detection performance were examined by comparing the distance at which targets were detected (with greater distance indicating better performance) and the detection accuracy [1-P(target missed)]. The data analysis also examined the effect of target expectancy by comparing the detection performance for the tanks, soldiers, and land mines – all highly expected targets – with the detection of the nuclear devices – infrequent, low expectancy targets. The data for the tanks and the soldiers, both expected and highly salient objects, showed similar trends; consequently, their data were collapsed for the analyses.

The selection of a hypothesis (in this case, whether or not a target was present) is often biased towards the most frequent and expected events. Hence, from here on, we use the term *unexpected* to refer to the nuclear device event because of its low frequency of presentation (on 10% of all target trials) relative to the other targets (tanks, soldiers, and land mines were presented each on 30% of the target trials). Although expectancy is confounded with physical differences between the stimuli, as in previous studies, contrast adjustments were made to ensure that the unexpected targets (nuclear devices) were *more* salient than the salient expected targets with which they were compared (tanks and soldiers). The nuclear device trials were separated into two classes based on whether the nuclear device was presented concurrently with a cued soldier or uncued soldier. Note that the presentation of the nuclear device only occurred with the soldiers.

The data were analyzed using a 2 (interactivity) x 2 (reliability) between subjects x 2 (scene detail) x 2 (cueing) x 3 [target type: expected, high salience (tanks, soldiers); expected, low salience (land mines); unexpected (nuclear device)] within subjects ANOVA. Note that the accuracy data refers to missed targets, not false classifications of distractors (on the 25% unreliable cueing trials).

Figure 7.1.1 shows the effects of cueing and target type for the target detection task.

(a) (b)

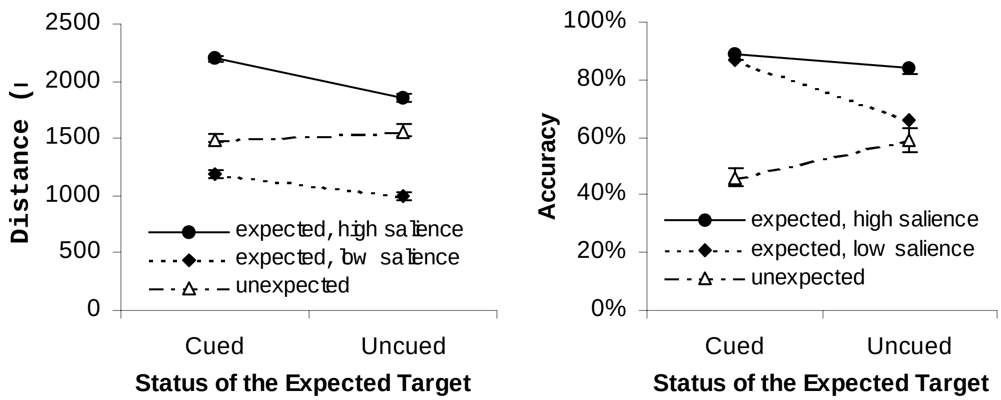


Figure 7.1.1. (a) detection distance and (b) accuracy as a function of cueing. Note that the unexpected event was never cued directly; instead the data presented in the figure are grouped according to the status of the object with which the nuclear device was paired.

Not surprisingly, the data analysis revealed a significant effect of target type on detection distance, $F(2,24)=113.41$, $p=0.0001$, and accuracy, $F(2,24)=94.05$, $p=0.0001$. Highly salient, expected targets were detected at greater distances and with higher accuracy than low salience targets. A significant cueing x target type interaction was also observed, $F(2,24)=4.09$, $p=0.03$. The results showed that the presence of cueing significantly increased the detection distance for the expected objects and did so strongly for the salient objects (tanks and soldiers), $F(1,12)=7.62$, $p=0.02$, and marginally increased the detection distance (improved performance) for expected but low salient objects (land mines), $F(1,12)=3.16$, $p=0.10$. The presence of cueing did not have an effect on the detection distance of the unexpected objects (nuclear device), $F(1,12)=0.76$, $p=0.40$.

As shown in Figure 7.1.1(b), the cueing x target type interaction, $F(2,24)=23.14$, $p=0.0004$, were manifest differently in terms of the accuracy data. While the presentation of cueing did not benefit the detection of expected, high salient targets, $F(1,12)=0.49$, $p=0.50$, it did significantly improve the detection of low salient objects, $F(1,12)=42.86$, $p=0.0001$. More important is the effect of cueing the soldier on the detection of the unexpected object – the nuclear device in the same scene. Subjects were instructed that the presence of the nuclear device was a high priority event, and if it was present, it needed to be detected *before* the object with which it was presented. Calculation of the accuracy data for the nuclear device detection thus took into account the order in which it was detected with respect to the soldier with which it was presented:

0 if it was not detected,

1 if it was detected and reported prior to detecting the soldier, and

0.5 if it was detected but reported after detecting the soldier with which it was presented.

The accuracy data revealed a significant effect of cueing, $\underline{F}(1,12)=6.72$, $\underline{p}=0.02$, in a direction opposite the trend present for the expected objects. That is, replicating effects we have observed in three earlier experiments (Merlo et al., 1999; Yeh et al., 1999, Experiments 1 and 2), the unexpected target was *less* likely to be detected when it was presented with a cued object (soldier) than an uncued one. The results confirm that detection behavior when cueing was present was influenced by attentional tunneling such that attention was allocated to the information presented by the cue, and subjects were less likely to scan around the cued area in order to detect a high priority but unexpected event.

The modulating effects of cue reliability on cueing benefits will be discussed in Section 7.4.

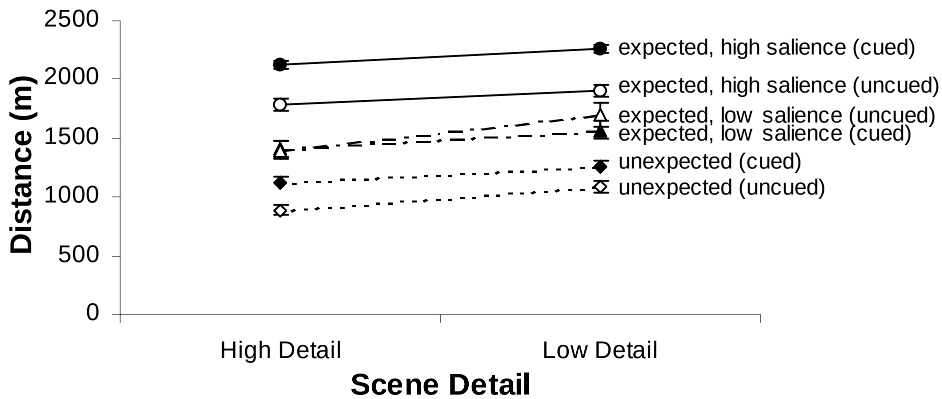
7.2 Image Realism

A comparison was made to assess whether or not the level of realism with which the scene was presented would influence the amount of attention allocated to the automated cueing information. Figure 7.2.1 presents the effects of scene detail and cueing on target detection distance and accuracy.

An examination of how the scene detail influenced the target detection performance data suggests an improvement in detection in low detail scenes. This advantage is interpreted in terms of the contrast and visibility of the targets in the scene relative to the texturing of the background [detection distance: $\underline{F}(1,12)=6.73$, $\underline{p}=0.02$; detection accuracy: $\underline{F}(1,12)=82.80$, $\underline{p}=0.0001$. The examination of image realism is inherently confounded with target detectability since differences in the texturing of the display, designed to create realism, influence the contrast between the object presented in the scene and its background, and hence the object's visibility. Although several steps were taken in pilot work to minimize the differences, a simple examination of target detection distance and detection accuracy for the uncued objects (open symbols in Figure 7.2.1) suggests that a difference was still present, and that this influence was greatest for objects of low salience. That is, land mines were detected at a greater distance, $\underline{F}(1,12)=4.48$, $\underline{p}=0.06$, and with significantly greater accuracy, $\underline{F}(1,12)=107.30$, $\underline{p}=0.0001$, in the low detail terrain than in the high detail terrain. While there was no effect of image realism on detection distance for the more salient soldiers or nuclear devices when no cueing information was available, $\underline{F}(1,12)=2.44$, $\underline{p}=0.14$ and $\underline{F}(1,12)=2.34$, $\underline{p}=0.15$, respectively, detection accuracy for both of these objects was higher in the low detail terrain than in the high detail terrain, $\underline{F}(1,12)=6.48$, $\underline{p}=0.03$ and $\underline{F}(1,12)=9.46$, $\underline{p}=0.01$, respectively. However, were it simply the case that all uncued objects were more difficult to detect against the high detail scene, the data should also show that subjects were less sensitive to the presence of targets in the high detail terrain than in the low detail terrain. This is an issue we explore later through signal detection analysis (discussed in section 7.4.3), in which we show that the actual sensitivity is equivalent between low and high scene detail, and that the improved detection rate in low detail is the result of a criterion shift.

The data for detection distance revealed that the presentation of cueing symbology did not mediate the effects of scene detail, $\underline{F}(1,12)=1.83$, $\underline{p}=0.20$. Instead, the benefit for cueing in minimizing the detrimental effects of high scene detail was observed in the accuracy data, $\underline{F}(1,12)=11.80$, $\underline{p}=0.005$. As Figure 7.2.1(b) shows, cueing information (filled points) improved target detection in the high detail scene to levels equal to that in the low detail scene. However, this cueing benefit was a function of target expectancy, $\underline{F}(1,12)=6.27$, $\underline{p}=0.007$, such that the presentation of cueing information aided the detection of the expected objects in the high detail scene [high salience: $\underline{F}(1,12)=5.12$, $\underline{p}=0.04$; low salience: $\underline{F}(1,12)=35.47$, $\underline{p}=0.0001$] but not the detection of the unexpected objects, $\underline{F}(1,12)=0.03$, $\underline{p}=0.86$.

(a)



(b)

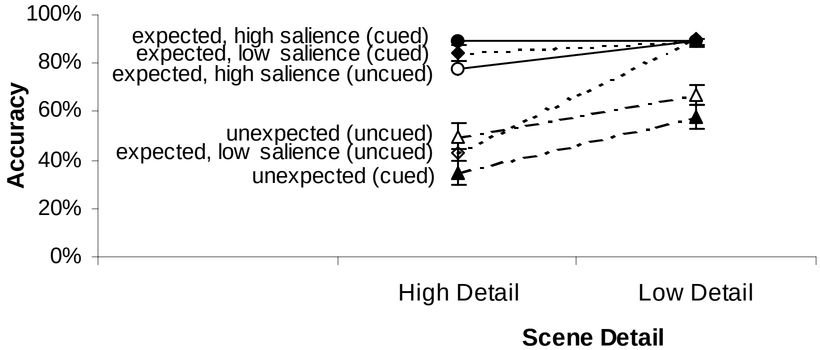


Figure 7.2.1. (a) detection distance and (b) detection accuracy as a function of cueing and image realism.

7.3 Interactivity

We had expected interactivity to produce attentional effects characterized by narrowing the focus of the region around the path and thus degrade the detection of peripheral versus central objects. In order to investigate this hypothesis, tanks and soldiers (expected, high salient objects) were presented in the periphery of the display and land mines (expected, low salient objects) were presented in the center, along the traveled path. To examine subjects' focus, the ImmersaDesk display was divided into nine areas of interest (top left, top center, top right, … bottom right) as shown in Table 7.3.1, and the amount of time subjects spent in each area, as reported by head movement data, was recorded. We had expected that active navigators would focus more on the tunnel pathway presented vertically along the center of the display, than the passive navigators, who would be more able to attend to information in the periphery. We found instead no difference in scanning (as inferred by head movement) across nine areas of interest. The results are presented in Table 7.3.1.

Table 7.3.1. The ImmersaDesk display was divided into 9 areas of interest (top left, top center, top right, … bottom right). The mean amount of time subjects spent in each of these regions are presented as a function of interactivity.

Active	1.56s	13.77s	1.79s
Passive	1.37s	14.40s	2.34s
	$F(1,12)=0.04$, $p=0.85$	$F(1,12)=0.10$, $p=0.75$	$F(1,12)=0.35$, $p=0.57$
Active	2.72s	42.73s	3.22s
Passive	3.33s	37.52s	3.06s
	$F(1,12)=1.17$, $p=0.30$	$F(1,12)=0.55$, $p=0.47$	$F(1,12)=0.09$, $p=0.77$
Active	0.34s	9.75s	0.52s
Passive	0.70s	11.17s	1.07s
	$F(1,12)=1.50$, $p=0.24$	$F(1,12)=0.03$, $p=0.87$	$F(1,12)=0.03$, $p=0.87$

A second analysis was conducted to compare the amount of time (as measured by head movements) spent in the center of the display relative to the periphery as a function of interactivity. For this analysis, the center was defined to be the center-most box of the 3x3 grid shown in Table 7.3.1; the periphery were the eight surrounding regions. The results showed no difference in the amount of time spent in the center versus the periphery due to interactivity, $F(1,14)=0.33$, $p=0.58$. However, the raw data suggested that active viewers spent about 9 seconds more in the center of the display (the location of the path they were following) versus the passive viewers who divided their attention relatively equally (less than a second difference) between the two areas. The analysis revealed that these differences were not significant, $F(1,11)=0.43$, $p=0.52$.

Consistent with these results, there was no salient main effect of interactivity on detection distance, $F(1,12)=0.72$, $p=0.41$, nor accuracy, $F(1,12)=0.13$, $p=0.72$, nor was there an interaction between interactivity and target type for either of these dependent variables [distance: $F(1,12)=0.84$, $p=0.44$; accuracy, $F(1,12)=1.89$, $p=0.17$]. Such an interaction would have been expected, had the passive condition availed more scanning, and hence more effective detection of the peripherally located tanks and soldiers.

Additionally, the analysis revealed no overall interaction between interactivity and cueing [distance: $F(1,12)=0.15$, $p=0.71$; accuracy: $F(1,12)=0.05$, $p=0.82$] nor was there an interaction between interactivity and image realism [distance: $F(1,12)=0.74$, $p=0.41$; accuracy: $F(1,12)=0.41$, $p=0.53$]. A 2 (interactivity) x 2 (reliability) between subjects x 2 (image realism) x 2 (cueing) within subjects ANOVA conducted on the data for each target type revealed a significant three-way interaction between scene detail, cueing, and interactivity for the distance with which land mines were detected, $F(1,12)=18.05$, $p=0.001$. This interaction is presented in Figure 7.3.1.

As Figure 7.3.1 shows, detection of the land mines in the low detail scene (as represented by the circles) benefited from cueing, and this benefit was independent of interactivity. In the high detail scene, however, land mine detection (as represented by the square icons) benefited from either active navigation or cueing, but not both. In other words, an active navigation mode only benefited the least detectable low salient mines in the high detail scene. The role of the attentional benefit and arousal benefit, shown in Figure 7.3.1, will be described in the discussion.

No other differences were significant within each target type.

However, an important role of interactivity was revealed in its interaction with reliability, which we discuss in the following section.

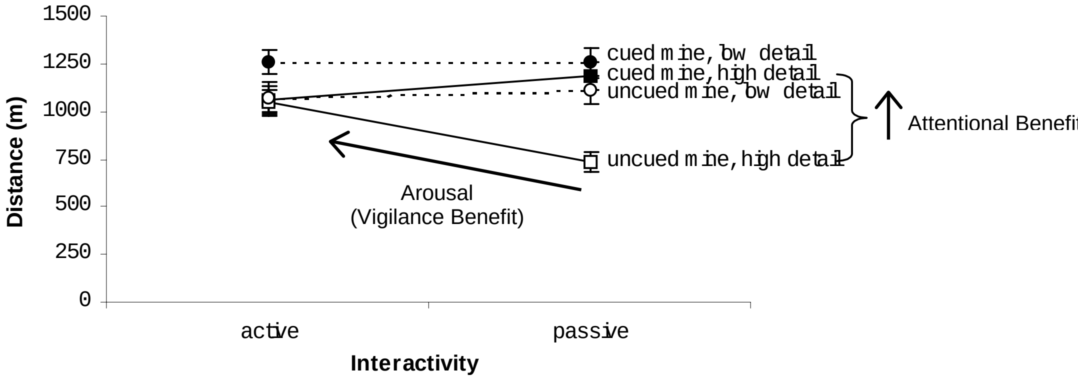


Figure 7.3.1. Land mine detection distance: scene detail x cueing x interactivity interaction. Note that these data are collapsed over the reliability manipulation.

7.4 Trust Effects and Biases: Cue Reliability

When cueing information is compelling, the reliance one places in that information, e.g., *overtrust*, may potentially have a “garden path effect”, leading one to an object that is not a target but which one believes (incorrectly) is a target. In operational circumstances, this can often happen when the automation that provides the cue falsely identifies a highly similar object (distractor) as a target. Here we operationally define high similarity by the truck-tank pair and the tree-soldier pair.

This trust in the cueing information when the cueing is less than reliable is expected to mediate attention allocation, and such behavior may be modulated by interactivity or scene realism. The manipulation of reliability is described in Table 7.4.1.

Table 7.4.1. Reliability manipulation.

Reliability Condition	Targets (tanks, soldiers, land mines)	Distractors (trucks/trees)
100%	Tanks always cued; soldiers and land mines cued 50% of the time	Never cued
75%	Tanks always cued; soldiers and land mines cued 50% of the time	Always cued (false alarm rate)

As Table 7.4.1 shows, the manipulation of reliability was implemented through the cueing of distractors; the presentation of target cueing was identical for the two reliability conditions: cueing information was presented for the tanks, half the soldiers, and half the land mines. When cueing was 100% reliable, distractors (trucks and trees) were never cued; when the cueing was only 75% reliable, however, of all the cues presented, 75% of these were targets and 25% of these were distractors, but every truck or tree that was present was cued. That is, all distractors were always cued, and the cueing of these objects equaled to 25% invalid trials.

Note that subjects in the partially reliable cueing condition were first presented with three blocks of 100% reliable cueing in order to foster their trust in the cue. The data from these trials were not included in the analyses for either group.

In section 7.4.1, trust and complacency effects are examined by comparing the false alarm rate, i.e., the number of detection of distractors when the object was cued in the 75% reliable condition (which would be characteristic of an automation overtrust) to those instances when the object was uncued as in the 100% reliable condition. The data are then examined to determine whether reliability influenced trust in the cue as observed in target detection performance for expected targets (section 7.4.2) and as reflected by changes in operator sensitivity and response criterion (section 7.4.3). Finally, the influence of operator trust on attention allocation is observed in the detection of the unexpected targets (section 7.4.4).

7.4.1 False Alarm Rate: Detection of Distractor Objects

In order to determine whether subjects were "led down the garden path," an analysis was conducted to determine the likelihood that subjects would detect a non-target as a target based on the information provided by a cue. A 2 (reliability) x 2 (interactivity) x 2 (target type) repeated measures ANOVA was conducted on the false alarm rate for the distractor objects (truck and tree). In the 75% reliable condition, distractors (i.e., a truck or tree) were cued on 25% of the cued trials, and the implementation of this manipulation resulted in the cueing of all trucks and trees present in the scene. That is, as Table 7.4.1 shows, the examination of cueing was inherent in the comparison of cue reliability as all the distractors in the 75% reliable condition were cued whereas no distractors in the 100% reliable condition were cued. The false alarm rates are presented in Figure 7.4.1.1.

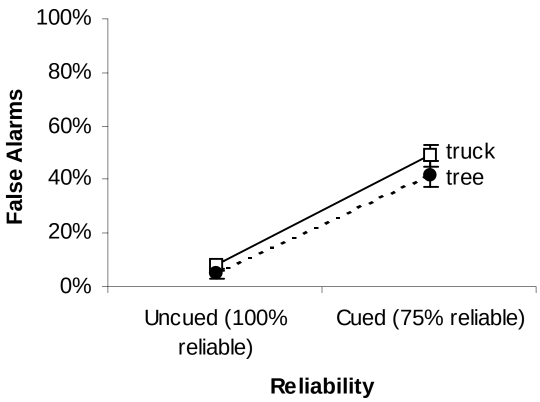


Figure 7.4.1.1. False alarm rate for distractor objects.

The analysis revealed a main effect of reliability, $F(1,12)=16.46$, $p=0.002$, such that distractor objects were detected as targets 46% of the time when they were cued but only 7% of the time when they were not, showing subjects' implicit trust in the cueing information, even when they knew it could be unreliable. The data also revealed a significant effect of target type, $F(1,12)=11.97$, $p=0.005$. The interaction between target type and reliability, $F(1,12)=5.01$, $p=0.05$, reflects the greater unreliability cost for the truck/tank pair, consistent with the view that for this more difficult discrimination pair, subjects became more reliant on the cue and suffered more when it failed (Wickens, Conejo, & Gempler, 1999). This difference is attributable to the greater similarity in appearance between the truck (distractor) and the tank (target) than between the tree (distractor) and the soldier (target).

A main effect of image realism revealed a higher rate of false detections (25%) in the low than in the high detail scene (21%), $F(1,12)=4.71$, $p=0.05$. A marginally significant interaction between reliability and image realism, $F(1,12)=3.80$, $p=0.07$, showed that this cost of low realism was present only when the cueing was 75% reliable. In this case, subjects were 51% likely to detect a cued false target in the low detail scene relative to only 41% in the high detail scene.

Additionally, the data revealed a main effect of interactivity such that subjects viewing the scene passively were more likely to detect false targets (32%) than active navigators (14%), $F(1,12)=5.12$, $p=0.04$. In other words, decreased interactivity apparently made subjects more complacent and consequently, more reliant on the cueing information, even when it was incorrect. Although statistics revealed that the interaction between interactivity and reliability was not significant, $F(1,12)=2.44$, $p=0.14$, the data presented in Figure 7.4.1.2 highlights the magnitude of the performance difference due to interactivity.

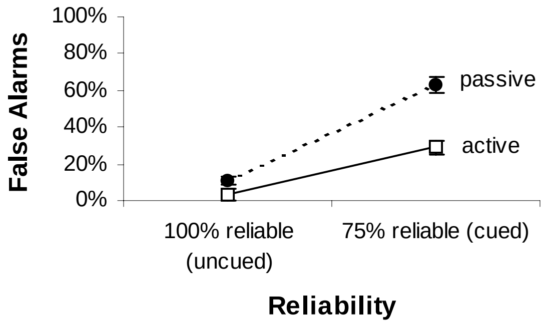


Figure 7.4.1.2. False alarm rate for distractor objects as a function of interactivity.

As shown in Figure 7.4.1.2, most of the false alarm cost in passive viewing occurred when the distractors were cued, i.e., in the partially reliable condition (a comparison of a false alarm rate of 46% when the distractors were cued versus only 7% when the distractors were not). This cost of unreliability, indicative of overtrust, was stronger in the passive condition (a 52% difference) than in the active condition (a 26% difference); these differences are substantial and suggest a strong trend that passive subjects showed greater overtrust of unreliable automation (more false alarms when it was unreliable). Thus, the data reveals that subjects did trust the unreliable cue and did so more when the difficulty of the task increased, e.g., when target-distractor discriminability was more difficult (i.e., distinguishing between a tank and a truck versus a tree and a soldier), or when they were more removed from the task, e.g., passively viewing the task.

7.4.2 Detection of Expected Targets

An analysis was then conducted to determine whether cue reliability influenced detection behavior (detection distance and accuracy) for expected objects, and how this behavior was modulated by interactivity and image realism. The data were analyzed using a 2 (reliability) x 2 (interactivity) between subjects x 2 (scene detail) x 2 (cueing) x 2 (target type: high salience, low salience) within subjects repeated measures ANOVA. The results are presented in Figures 7.4.2.1 - 7.4.2.3.

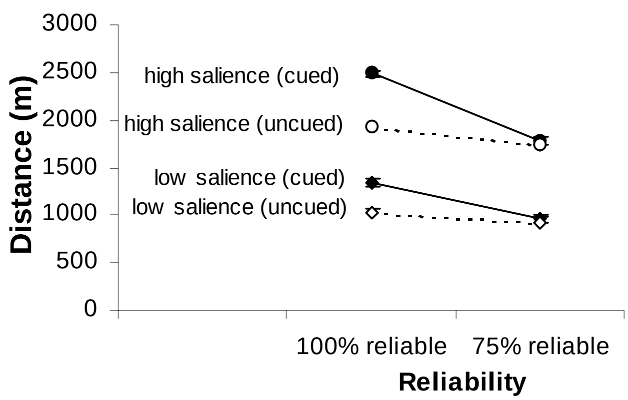


Figure 7.4.2.1. Effects of reliability as a function of cueing and target salience.

As shown in Figure 7.4.2.1, unreliability degrades performance, $F(1,12)=10.64$, $p=0.01$, but does so exclusively on cued trials, $F(1,12)=5.69$, $p=0.03$ (there was no significant effect of reliability on uncued trials, $F(1,12)=1.04$, $p=0.33$). That is, subjects distrust the cue more and do not use it to their advantage when it has proven to be less than perfectly reliable. These data reveal that the "distrust" in the cue when it is unreliable is mediated by interactivity, as shown by a significant interaction between cue reliability and interactivity, $F(1,12)=8.21$, $p=0.01$, which revealed soldiers who were actively navigating were hurt more by the presence of system unreliability than those engaged in passive navigation. This is shown in Figure 7.4.2.2.

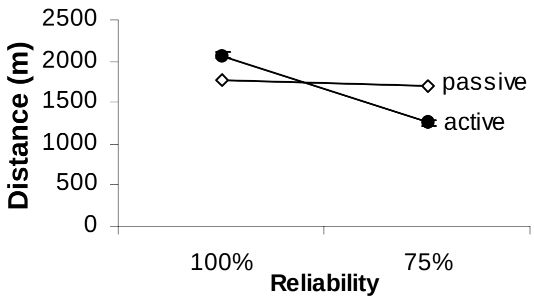


Figure 7.4.2.2. Target detection distance as a function of interactivity and cue reliability.

The data in Figure 7.4.2.1 suggests the presence of a three-way interaction, such that the 100% reliable cueing benefits were greater for the high salience than the low salience targets, and although this interaction was not significant, $F(1,12)=1.81$, $p=0.20$, the data does reveal the presence of a strong trend confirming the effect.

Instead, the apparent three-way interaction is manifest in a significant four-way interaction involving control interactivity, $F(1,12)=5.02$, $p=0.04$, which may be seen in Figure 7.4.2.3 in which the same graph is plotted for the active group in the top panel and the passive group in the bottom panel.

The data shown in Figure 7.4.2.3 suggests that the automation-failure reduction in cue benefits was modified by both activity and salience. That is, only when the target was highly salient and the subject was actively navigating through the scene was the maximum cost of unreliable target cueing imposed.

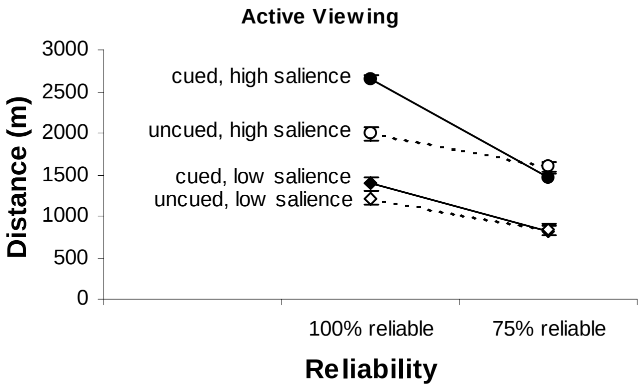
Note that none of the other three-way interactions approached significance: reliability x cueing x interactivity: $F(1,12)=0.54$, $p=0.48$; cueing x interactivity x saliency: $F(1,12)=0.23$, $p=0.64$.

Thus, the target detection distance data suggests that active navigators were more resource-limited when confronted with potentially unreliable cueing information and these resource limits hindered their detection of high salience targets. Consequently, we might also expect that performance on other tasks being performed simultaneously should also suffer due to limitations in resource demand and allocation. That is, the data should show not only decrements in target detection distance but also decrements in control precision. To test this hypothesis, a 2 (reliability) x 2 (image realism) x 5 (path) repeated measures ANOVA was conducted on the data describing the active navigator's control deviations from the specified path (passive viewers performance is by definition automatically controlled and hence cannot be included in the analysis). The results revealed performance decrements when the cue was only partially reliable as shown by a significant 72% increase in RMSE, $F(1,6)=9.13$, $p=0.02$, when the cueing was 75% reliable relative to when it was 100% reliable.

The results of the literature review suggested that as image quality decreased, subjects would be less likely to trust the cueing information (MacMillan et al., 1994). Consequently, we hypothesized that as the level of scene realism decreased, subjects would trust the cue less, scan the environment more, and be more cautious about reporting a target on the basis of a cue, i.e., decreased target detection distance. The data did not confirm this hypothesis; the interaction between cue reliability and image realism was not significant, $F(1,12)=1.09$, $p=0.32$.

Detection accuracy as a function of cue reliability is presented in Figure 7.4.2.4.

(a)



(b)

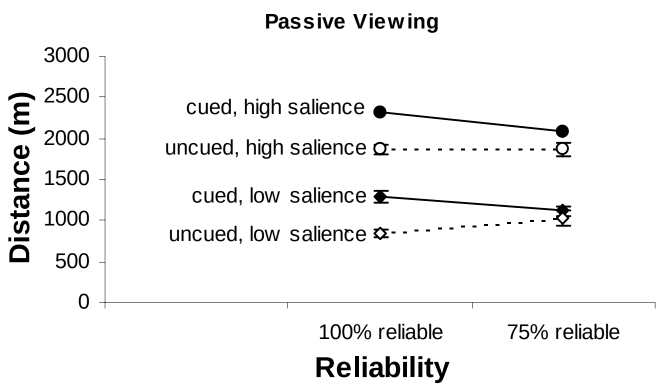


Figure 7.4.2.3. Target detection distance as a function of cueing and cue reliability for (a) active viewing and (b) passive viewing.

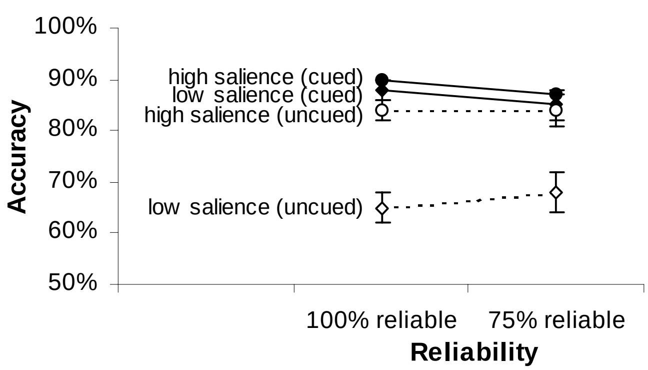


Figure 7.4.2.4. Detection accuracy as a function of cue reliability.

The accuracy data revealed no effects of cue reliability on overall target detection, $F(1,12)=0.01$, $p=0.94$, nor were there significant interactions between cueing and reliability, $F(1,12)=1.77$, $p=0.21$, interactivity and reliability, $F(1,12)=1.08$, $p=0.32$, nor detail and reliability, $F(1,12)=0.00$, $p=0.97$.

7.4.3 Signal Detection Analysis

Target detection in the scenario used for the current experiment required that the subject discriminate between two discrete object states, i.e., whether the object that was present was a target or not. This discrimination task was difficult and errors were made (as reflected by the false alarm rate in Figure 7.4.1.2). To determine whether the nature of these errors were attributable to a lack of sensitivity in the target discriminability task or simply an inclination to respond that an object was a target, signal detection analysis was conducted on these data in order to determine how automated attentional guidance combined with different features of realism (i.e., image realism and interactivity) could modulate the soldier's sensitivity and the setting of a response criterion. Due to the similarity in the saliency (i.e., shape and size) of the tank to the truck and the soldier to the tree, only the data from those four objects were used for the analysis. The results are reported below.

Sensitivity was estimated using the non-parametric measure,

$$P(A) = \frac{P(H) + [1 - P(FA)]}{2}$$

where *H* is the hit rate based on the detection accuracy for the highly salient and expected object, and FA is the rate of false alarms (Wickens, 1992).

ß was estimated using the formula:

$$ß = e\hat{}\{-[z(H)^2 - z(1-FA)^2]/2\}$$

where *z* is the z-score, *H* is the hit rate based on the detection accuracy, and *FA* is the rate of false alarms (Procter & van Zandt, 1994).

In order to assess how reliability influenced reliance and/or trust in the cueing symbology, P(A) and ß were calculated for uncued events, for cued events when the cueing was 100% reliable, and for cued events when the cueing was only partially (75%) reliable. An examination of cueing x reliability was constrained by the fact that the manipulation of reliability was implemented simply by the cueing (or non-cueing) of distractors. This is shown in Table 7.4.3.1.

Table 7.4.3.1. Availability of hit rates and false alarms for cued and uncued targets as a function of reliability.

	100% reliable			**75% reliable**	
Cueing	**HR**	**FA**		**HR**	**FA**
Uncued	✓	✓		✓	----
Cued	✓	----		✓	✓

As Table 7.4.3.1 shows, when cueing information was 100% reliable, *all* the distractors that were presented were *uncued*; consequently, a false alarm rate for cued objects in the 100% reliable condition could not be calculated directly (i.e., there were literally no opportunities to make false alarms). On the other hand, when cueing information was only partially reliable, all the distractors were *cued*; consequently, a false alarm rate for uncued objects and distractors could not be calculated with these data. Thus, the calculations of P(A) and ß for uncued events (a baseline of subjects' sensitivity and response criterion when no cueing was available) were based solely on the data from subjects in the 100% condition.

We wished to estimate how subjects' sensitivity and response criterion would change in the detection of *cued* objects when the guidance information became unreliable – a comparison of reliance when cueing was completely reliable versus only partially reliable. However, this comparison could not be made directly given the current design. In order to make such a comparison, P(A) and ß were estimated using the data from the first 3 blocks of the partially unreliable condition (i.e., those blocks when subjects in the 75% reliable cueing condition were presented with 100% accurate cueing to foster trust in the attentional guidance) to calculate the hit rate supported by 100% reliable cues. The false alarm rate for the 100% reliable cueing was estimated from those subjects' response to the first cued distractor in the fourth block, i.e., the first block in which the cueing "failed" and guided attention to a distractor. Since this event occurred only once per participant, a single probability estimate of this value was obtained for all subjects. The results for P(A) and ß as a function of cueing and reliability are presented in Table 7.4.3.2.

Table 7.4.3.2. Signal detection results: cueing x reliability.

Condition	P(hit)	P(FA)
Uncued	(# targets reported) / # targets seen 84%	# distractors reported (false targets) / #nontargets 8%
Cued (100% reliable)	# cued targets detected / total cued targets 91%	# cued distractors reported (as targets) / #nontargets (first trial) 63%
Cued (75% reliable)	# cued target detected / total cued targets 89%	# cued distractors reported (as targets) / total cued targets (subsequent trials) 45.5%

Uncued	P(A)=0.88, ß=2.04
Cued (100% reliable)	P(A)=0.64, ß=0.43
Cued (75% reliable)	P(A)=0.71, ß=0.77

As Table 7.4.3.2 shows, subjects were generally less sensitive when cueing symbology was available to aid them in the target detection task. More importantly, as subjects' sensitivity decreased (i.e., their reliance on the guidance *increased*), their response criterion shifted so that responses were riskier when they believed that the cueing information was reliable. That is, subjects trusted the information suggested by the cue rather than examining the raw data underneath, an effect replicating that found in a different context by Wickens et al. (1999). However, when subjects were presented with repeated instances of the automation failure (following block 4), their sensitivity (and trust in the system) was re-calibrated. Sensitivity improved, but not to the level originally seen with no cueing whatsoever. Their response criterion also adjusted to show a reduced willingness to report a target. However, they were still somewhat guided by advice of the cue, as witnessed by the riskier beta setting than in the uncued condition.

While a direct comparison of image reality as measured by target detection performance was confounded by differences in the texturing of the display, signal detection analyses provided a measure of reliance on the image itself as well as the cue through an examination of differences in sensitivity and ß shifts in the high detail terrain relative to the low detail terrain. If differences in the target-background contrast were in fact responsible for the performance differences, one would expect a corresponding sensitivity difference, with a decreased sensitivity in the high detail scene relative to the low detail one. However, no such difference was found (as shown in Table 7.4.3.3 which collapses data across the two cueing conditions) but rather the data reveals a greater willingness to report a target (a lower ß) in the low detail scene.

Table 7.4.3.3. Signal detection results: scene detail.

Realism	HR	FA	
High Detail	86%	22%	P(A)=0.81, ß=1.33
Low Detail	89%	26%	P(A)=0.79, ß=0.81

More interesting was the effect of scene realism on reliance on cueing information, specifically when the cueing information failed the first time. This effect is shown in Figure 7.4.3.1.

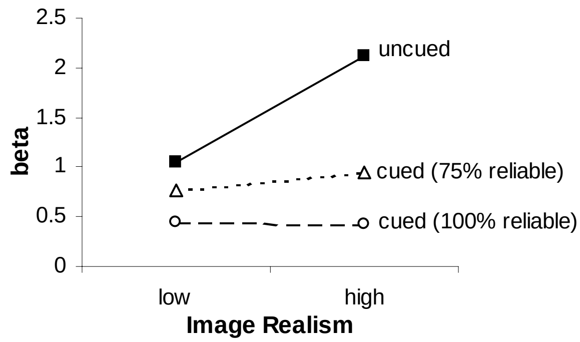


Figure 7.4.3.1. Shifts in ß as a function of realism and cueing.

The figure suggests that raw data in the high detail scene were processed to a greater extent (e.g., higher ß) as the informativeness of the attentional guidance decreased. The data also suggest that cue availability mediated the effect of realism on the willingness to report a target. In examining Figure 7.4.3.1, the response criterion changed little with realism when targets were cued reliably (the bottom line), but as the cueing information became less reliable (75%) or less available (uncued), the data reveal a progressive trend towards a more conservative bias, particularly with a highly realistic scene. If the difference in criterion between uncued and cued conditions can be assumed to reflect an implicit measure of cue reliance, then we see that such reliance is enhanced when realism is high.

While increased realism resulted in riskier responding (a shift in ß), increasing the level of control through active navigation resulted in increased sensitivity. The results are shown in Table 7.4.3.4.

Table 7.4.3.4. Signal detection results: interactivity.

Interactivity	HR	FA	
Active	86%	15%	P(A)=0.84, ß=1.47
Passive	89%	32%	P(A)=0.76, ß=0.97

The analysis revealed that active navigators were more sensitive and less complacent in their responses. Passive viewers compensated for the loss in sensitivity by a downward beta shift to preserve a high hit rate.

7.4.4 Detection of Unexpected Objects

If subjects showed biases in trust and attention, then the presence of a highly reliable cue should capture attention at the expense of the higher priority unexpected objects present in the scene. Consequently, detection of the unexpected objects would be expected to be higher when cueing was only 75% reliable than when cueing was 100% reliable. Such a result was reported by Merlo et al. (1999). In order to test this hypothesis, a 2 (reliability) x 2 (interactivity) between subjects x 2 (cueing) within subjects repeated measures ANOVA was conducted on the detection distance and accuracy for the unexpected object (the nuclear device). The results of the analysis are presented in Figure 7.4.4.1.

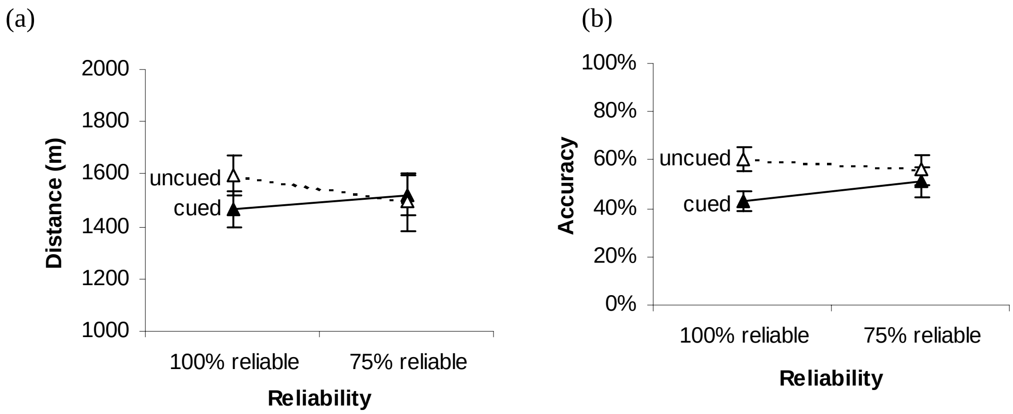


Figure 7.4.4.1. Attention and Trust Biases: Effects of cue reliability on (a) the detection distance and (b) detection accuracy of the unexpected objects (nuclear devices). Note that the unexpected event was never cued directly; instead the data presented in the figure are grouped according to the status of the object with which the nuclear device was paired.

The results for detection distance show no interaction between the presentation of cueing and its reliability, $\underline{F}(1,12)=0.07$, $\underline{p}=0.79$, nor did the accuracy data reveal this bias, $\underline{F}(1,12)=0.87$, $\underline{p}=0.37$. However, due to the low power of this test (a high variability and few observations of the nuclear device detection) and the strong visual pattern in Figure 7.4.4.1(b), we examined the two reliability conditions separately for the accuracy of nuclear device detections. When the data was analyzed in this way, the accuracy data for the 100% reliable trials revealed a significant cost to cueing, $\underline{F}(1,6)=7.67$, $\underline{p}=0.03$, whereas the data for the 75% reliable trials did not, $\underline{F}(1,6)=1.16$, $\underline{p}=0.32$, as clearly suggested by the overlapping standard error bars on the right side of Figure 7.4.4.1(b).

The data for detection distance were analyzed in this same way but revealed no difference [100% reliable: $\underline{F}(1,6)=0.69$, $\underline{p}=0.44$; 75% reliable: $\underline{F}(1,6)=0.17$, $\underline{p}=0.69$].

7.5 Trust Effects due to Image Realism: Checkpoint Detection

We had hypothesized that a higher trust would result from greater image realism, so that as scene fidelity increased, subjects would trust the information in the simulation, which they were warned could be incorrect, even when such trust was not desirable. To test this hypothesis, subjects were given a terrain association task, which required them to call out checkpoints as they were passed. The checkpoints consisted of salient terrain features and were marked on a hand held paper map.

In order to evaluate the trust effects attributable to image realism, a 2 (interactivity) x 2 (reliability) between subjects x 2 (image realism) within subjects repeated measures ANOVA was conducted on the detection distance error and accuracy data for checkpoint detection. Subjects viewing the high detail scene detected 63% of the checkpoints at a distance of 501m; in the low detail scene, subjects detected 73% of the checkpoints at an average of 505m. These differences were not significant [distance: $\underline{F}(1,12)=0.05$, $\underline{p}=0.83$; accuracy: $\underline{F}(1,12)=2.88$, $\underline{p}=0.12$].

There was also no effect of interactivity [distance error: $\underline{F}(1,12)=0.12$, $\underline{p}=0.73$; accuracy: $\underline{F}(1,12)=0.01$, $\underline{p}=0.94$] nor reliability [distance error: $\underline{F}(1,12)=0.04$, $\underline{p}=0.85$; accuracy: $\underline{F}(1,12)=0.00$, $\underline{p}=0.98$], nor were the interactions between these variable significant [distance error: $\underline{F}(1,11)=1.61$, $\underline{p}=0.23$; accuracy: $\underline{F}(1,11)=0.01$, $\underline{p}=0.94$].

In order to implicitly measure trust, subjects were asked to report any inconsistencies between the information in the map and the computer generated imagery. In the last block of trials, one mountain, located next to a checkpoint, was removed from the simulation. If greater image realism did indeed foster higher trust, then subjects would be more likely to detect the missing map feature in the low detail terrain than in the high detail terrain.

None of our subjects explicitly noticed the missing mountain. Two subjects (one who participated in a pilot study) may have implicitly noticed this error by failing to call out the checkpoint. When asked, they commented that although the cue they had been searching for had been the missing mountain (which they did not see), their failure to detect the checkpoint could also be attributable to the fact that the missing mountain occurred at a high workload moment. That is, at that point, subjects were detecting not only a target (a soldier), but also a high priority target (the nuclear device), and active navigators were in the middle of a 90° right turn.

7.6 Subjective Measures

Three categories of subjective measures were collected:

7.6.1 Workload: NASA-TLX Ratings

7.6.2 Subjective assessments of image realism

7.6.3 Subjective assessments of trust

Each of these ratings were analyzed using a 2 (interactivity) x 2 (reliability) between subjects x 2 (image realism) within subjects ANOVA.

7.6.1 Workload: NASA-TLX Ratings

The results of the workload ratings as a function of interactivity as measured by the NASA-TLX index are presented in Figure 7.6.1.1.

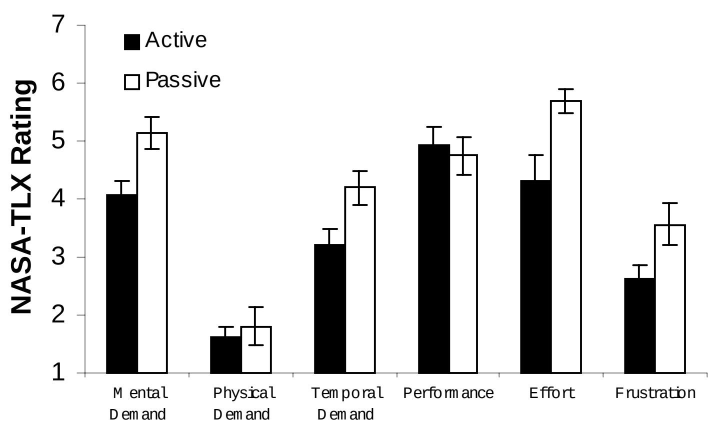


Figure 7.6.1.1. NASA-TLX Ratings. For each of these scales, a rating of 1 signified low workload and a rating of 7 signified high workload.

Subjects reported marginally higher mental demand, $F(1,12)=4.45$, $p=0.06$, temporal demand, $F(1,12)=3.28$, $p=0.10$, and effort, $F(1,12)=3.80$, $p=0.08$, when they passively viewed the simulation. There was no difference in the ratings as a function of interactivity for physical demand, $F(1,12)=0.15$, $p=0.70$, performance, $F(1,12)=0.22$, $p=0.65$, nor frustration, $F(1,12)=3.00$, $p=0.11$.

7.6.2 Subjective Assessments of Image Realism and Trust

Subjects were asked to rate the compellingness of the imagery and to assess the level of trust they placed in the terrain information. Their responses are presented in Table 7.6.2.1.

Table 7.6.2.1. Subjective ratings of the compellingness of the terrain imagery and the degree of trust they placed in the simulation.

	Image Realism	High Detail	Low Detail	
R1	How much did the visual aspects of the environment involve you? (1 = not at all, 7 = completely)	4.69 (0.36)	3.94 (0.42)	**F(1,12)=6.55, p=0.03**
R2	How compelling was your sense of moving around inside the virtual environment? (1 = not compelling, 7 = very compelling)	5.44 (0.27)	4.94 (0.34)	F(1,12)=2.74, p=0.12
R3	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities? (1 = not at all, 7 = prevented task performance)	3.31 (0.34)	2.81 (0.34)	F(1,12)=1.85, p=0.20
R4	Were there moments during the virtual environment experience when you felt completely focused on the task or environment? (1 = none, 7 = frequently)	5.25 (0.31)	5.38 (0.41)	F(1,12)=0.11, p=0.75
R5	What was your overall enjoyment in navigating through the environment? (1 = highly enjoyable, 7 = not enjoyable at all)	2.88 (0.38)	2.75 (0.41)	F(1,12)=0.09, p=0.77
R6	How would you rate the fidelity (i.e., the goodness) of the display? (1 = not realistic at all, 7 = very realistic)	5.38 (0.18)	4.50 (0.37)	**F(1,12)=8.40, p=0.01**
T1	How much did you trust the terrain information presented in the computer generated imagery? (1= not at all, 7 = very much)	4.69 (0.25)	4.19 (0.29)	F(1,12)=2.18, p=0.17
T2	How much did you trust the information presented in the paper map? (1= not at all, 7 = very much)	5.81 (0.33)	5.75 (0.40)	F(1,12)=0.10, p=0.75
T3	How much did you trust the cueing information? (1 = not at all, 7 = very much)	4.50 (0.45)	5.19 (0.31)	**F(1,12)=8.07, p=0.01**
T4	How high was your confidence in your ability to locate targets? (1 = low, 7 = very confident)	5.31 (0.25)	4.75 (0.30)	**F(1,12)=3.42, p=0.09**

As Table 7.6.2.1 shows, the difference in realism between the two terrain images was perceived both explicitly (question R6) and implicitly because subjects found the high detail imagery more compelling (question R1), but this compellingness did not influence the degree of trust they placed in the system (question T1), nor in the cues that were presented in the context of the images (question T3). While the data suggest that subjects trusted the cue *less* in the high detail scene than in the low detail one (question T3), the presence of a significant scene detail x cue reliability interaction, $\underline{F}(1,12)=8.07$, $\underline{p}=0.01$, reveals that this distrust of the cue in the high detail scene occurred primarily when the cue was only partially reliable. That is, the ratings revealed equal levels of trust in both high and low scene detail levels when the cueing information was 100% reliable (ratings of 6.0), but significantly *less* trust in the cue in the high detail scene (rating of 3.0) relative to the low detail scene (rating of 4.38) when the automated guidance was only 75% reliable, $\underline{F}(1,6)=9.31$, $\underline{p}=0.02$. This distrust of automation in the high detail scene confirm the signal detection results reported in section 7.4.3, which revealed a more conservative response criterion in the high detail scene relative to the low detail scene as the cue became less informative. It also follows then that subjects would be more confident in their ability to detect targets in the high detail scene than in the low detail one (question T4).

Not surprisingly, when subjects were asked to rate the degree of trust they placed in the cueing information, a main effect of cue reliability was present, such that subjects were more likely to trust the cue when it was 100% reliable (rating of 6.00) than when it was only 75% reliable (rating of 3.69), $F(1,12)=35.10$, $p=0.0001$. What is surprising though was an effect of image realism in subjects' response, revealing that they trusted the cue more when it was presented in the low detail scene (rating of 5.19) than when it was presented in the high detail scene (rating of 4.50), $F(1,12)=8.07$, $p=0.01$.

8. DISCUSSION

8.1 Overview

The current experiment was conducted to determine whether manipulations of image reality, interactivity, and cue reliability would influence attention allocation and trust attribution to different sources of information within an augmented reality environment. The presentation of each of these display features is compelling, naturally capturing attention and modulating behavior in such a way that one could fail to notice critical items in a scene (e.g., a high priority but unexpected target) or fail to question the validity of the data when it is critical to do so (i.e., overtrust as evidenced by the likelihood of missing errors in the terrain database or detecting a distractor as a target).

Figure 8.1.1 presents a causal model describing how the display features manipulated in the current study were predicted to influence the attention deployment throughout the visual scene and the trust of different information sources. The four display features (cueing, interactivity, reliability, and realism) are presented at the left; the measures used to assess their costs and benefits are listed to the right.

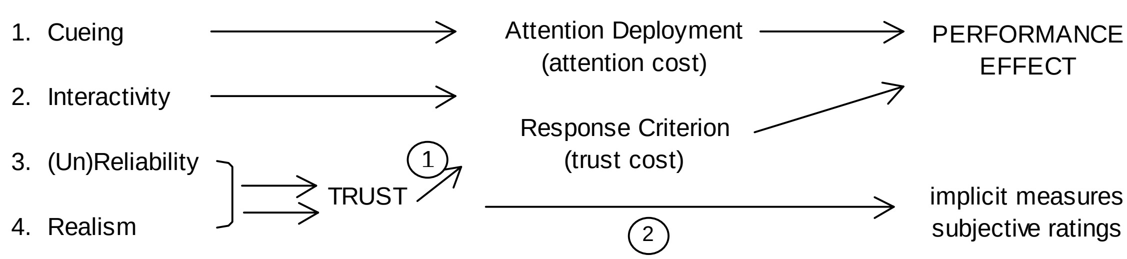


Figure 8.1.1. Causal model: cueing, image realism, reliability, and interactivity influences on attention allocation and trust effects as measured by explicit (performance effect) and implicit (subjective ratings) measures.

As shown in Figure 8.1.1, *cueing* and *interactivity* were expected to directly influence *attention deployment* throughout the scene: the former by directing attention to the location of cued targets camouflaged in the scene and the latter to those aspects of the simulation relevant to the manual control task, e.g., the tunnel pathway superimposed over the terrain. The presentation of cueing information was expected to modulate attention allocation behavior in such a way that attention would be allocated to the region highlighted by the cue and less to the rest of the environment (Merlo et al., 1999; Ockerman & Pritchett, 1998; Yeh et al., 1999), consequently either influencing subjects' *response criterion* (willingness to report a target) or sensitivity (closer inspection of the raw data at the cued location) as measured by target detection performance (expected hits vs. false alarms). While interactivity was not hypothesized to directly influence trust in the cueing information, the increase in workload imposed by the active control task was expected to increase subjects' reliance on the cueing information (McFadden et al., 1998; Riley, 1994).

Reliability of the cueing information and the *realism* of the scene were expected to induce miscalibrations of *trust* in the information presented by the cue or simulation data, respectively, in such a way that would also modulate *attention deployment* in a predictable way (as shown by arrow 1) but their influence was also measured implicitly and subjectively (as shown by arrow 2).

It was expected that operators would become dependent on the automation to inform them as to the location of targets in the environment, failing to detect errors when they occurred, i.e., *automation-induced complacency* (Parasuraman & Riley, 1997). This miscalibration of trust in the automated guidance information was expected to result in a cost-benefit relationship for the presentation of cueing (similar to that shown by Posner & Snyder, 1975). That is, benefits were expected when a cue was highly reliable and/or correct with costs to performance (e.g., an increase in false alarms as found by Conejo & Wickens, 1997) when the cue was less reliable or wrong.

Enhancing *image reality* could potentially increase operator confidence in the displayed data to a level that was inappropriate given the integrity of the data (Theunissen, 1998). Additionally, previous research by MacMillan et al. (1994) suggested that trust in automated guidance information could be influenced by the image quality: operators relied more on the visual image of the object rather than the automation's advice when the image quality is poor. This finding was consistent with the idea of automation *undertrust* (Lee & Moray, 1994; Parasuraman & Riley, 1997) with implications for simulator design such that a highly realistic scene with its higher image quality could induce greater trust in the cue relative to the low detail scene. These assessments of trust were measured <u>explicitly</u>, through target detection performance behavior (detection of expected versus unexpected high priority events) and subjective ratings of trust and perceived realism, and <u>implicitly</u>, by requiring subjects to detect an error in the simulation database (a missing mountain) based on information presented on a less compelling paper map.

The data are partially consistent with this causal model shown in Figure 8.1.1 but not totally, revealing some of the links are not as strong as was predicted and some other <u>unanticipated</u> mediating variables (e.g., visibility and resource demand, not attention allocation). As an overview of the results, our data reveal that presentation of cueing information is compelling; subjects used the cue to their advantage when a target was cued, but failed to allocate attention optimally, especially when the objects to be detected were unexpected. This misallocation of attention was mediated somewhat by less reliable cueing, i.e., a reduced trust in the cue allowed subjects to more effectively allocate attention throughout the environment, although trust biases were still present such that subjects were more likely to detect a non-target object as a target when that object was cued than when it was not.

The two display enhancements to support users' interactions which were provided – increasing image reality and interactivity, did indeed modulate attention deployment but not in a direction we had predicted. Although subjects did find the high detail scene more compelling and more involving, this greater image realism did not increase trust in the simulation itself. However, the data reveals a criterion shift such that subjects were more reliant on the guidance information when it was presented in the high detail scene and less likely to process the raw data, particularly as the informativeness of the guidance information increased. Interaction with the simulation alone – to accomplish navigation – did not increase workload in such a way that subjects were more reliant on cueing information. In fact, it appeared to reduce reliance on the cue, as active subjects showed fewer false alarms when the cue became unreliable (Figure 7.4.1.2). More interesting though was an interaction between cue reliability and interactivity, such that active navigators had less resources to deal with the loss of reliability of the cueing information. The degree of this experimental interaction was dependent on the presentation of cueing information and target saliency.

Each of these effects will be discussed in greater detail below.

8.2 Cueing

The results replicate previous findings of cueing benefits and costs (Merlo et al., 1999; Yeh et al., 1999), but in a different paradigm involving active and continuous navigation. The presentation of cueing information was effective (particularly when it was reliable), and these benefits were enhanced when the object that was cued was low in salience, e.g., the accuracy of detecting land mines benefited more from the presentation of cueing than did the soldiers, and did so particularly when they were less visible against the high detail background.

However, there were two visible costs to the use of this automated attentional guidance. (1) The signal detection analysis reveals a shift in the response criterion when cueing was available without a corresponding change in sensitivity, so that subjects were riskier in their responses and likely to overlook errors in the automated cueing system, resulting in a higher false alarm rate. Cueing did not apparently lead them to examine the raw data at the cued location more carefully, an effect which would have been manifest as an increase in sensitivity. (2) The data reveals an attentional bias characterized by the decreased detections of the unexpected but high priority target when it was presented concurrently with a cued expected target. This attentional tunneling is one example as to the compellingness of automated cueing, in which automation implicitly captures attention and guides it elsewhere at the expense of other objects in the environment, a phenomenon which has been reported in tasks ranging from target detection (Merlo et al., 1999; Yeh et al., 1999) to aircraft maintenance fault inspection (Mosier et al., 1998; Ockerman & Pritchett, 1998). This degree of this attentional tunneling was mediated by the reliability of the cueing information, an issue we now address.

8.3 Reliability

The cueing effects were mediated by the reliability of the cueing information. Unreliable cueing had three primary effects:

(1) reducing the benefits of cueing,

(2) increasing the rate of false alarms, and

(3) decreasing the attentional cost of cueing.

First, the results suggest that subjects trusted the cue more than was warranted. Such behavior was evidenced by overreliance on the cueing information, as evidenced by a higher rate of false alarms (i.e., the detection of false targets) when the cue guided attention to the location of a non-target, i.e., a truck or tree. Similar reports of overreliance on automated information led to subjects attacking non-target areas based on information suggested by a cue (Conejo & Wickens, 1997), avoiding non-targets that were cued (Davidson & Wickens, 1999), and focusing on what an aircraft was predicted to do (unreliably) in the future (Wickens, Gempler, & Morphew, 2000). The results of the signal detection analysis revealed that this overreliance was attributable to a criterion shift, similar to that reported by Swennsen et al. (1977), such that subjects were not only more likely to report the presence of a target object, but also the presence of a distractor.

Subjective ratings clearly reveal that subjects trusted the cueing information less when it was only 75% reliable, but their inclination nevertheless to attend to the cueing information rather than the underlying visual information (target shape) could be attributable to the two following factors. (1) Subjects were more confident in their own sensory abilities than that of the automated guidance system (a bias reported previously by Entin, 1998) so that they adjusted their criterion to a more conservative level when processing the raw data as the cueing information became less informative. (2) The target discriminability task was difficult; consequently, the use of automation increased, even when that automation was not reliable (as predicted by Lee & Moray, 1992). Consistent with this second interpretation, Gempler and Wickens (1998) reported that trust of unreliable predictor information was greatest on the most difficult trials.

A consequence of this unreliability was the reduction in benefits for cueing. As shown in Figure 7.4.2.1, detection distance was reduced when cueing was only partially reliable relative to when it was always reliable, and the performance cost for partially reliable cueing was such that the presentation of cueing for *valid* targets provided no benefits in terms of detection distance relative to uncued objects. In fact, the data revealed that when subjects knew that the automation was only partially reliable, they detected cued targets at reduced distances relative to those subjects presented with 100% reliable cueing (note that there was no difference in accuracy). This pattern of results is consistent with Posner and Snyder's (1975) cost-benefit relationship in target detection performance relative to cue reliability, i.e., a benefit for cueing when it was highly reliable which increased at a rate faster than the cost of presenting an unreliable cue. Similarly, Fisher et al. (1989) reported reduced benefits for highlighting key elements of a list on a display when that highlighting was less than fully reliable.

The distribution of attention was apparently changed by unreliability in such a way that unreliability *decreased* the attentional tunneling induced by cueing symbology. When an operator is given multiple sources of information, one information source may be favored at the expense of another (as shown by Conejo & Wickens, 1997), and the reliance that one places in that information may guide attention away from critical information in the scene (Merlo et al., 1999; Mosier et al., 1998; Ockerman & Pritchett, 1998; Yeh et al., 1999). The detection data reveals this behavior in the detection of the unexpected but high priority target, as shown in Figure 7.4.1.1; subjects focused on that region in space highlighted by the cue, even when it was not optimal to do so. However, when the cue was partially reliable, subjects' loss in trust moderated and broadened their visual search strategies in such a way that they were less reliant on the cueing information and more likely to scan the surrounding environment, detecting the unexpected, high priority events, as shown by the accuracy data in Figure 7.4.4.1. The findings replicate those reported by Merlo et al. (1999) in which subjects' breath of attention was widened when subjects' trust of the automation was betrayed.

The reliability effects were modified by other factors. In particular, the presence of active viewing coupled with unreliablilty appeared to lead to less use of the cues (fewer false alarms, less benefits of cued targets), but more difficulty overall (fewer detections of <u>all</u> targets, including those that were not cued) and poorer tracking performance. The behavioral data suggests that the loss of reliability imposed more resource demand on visual search and detection, which increased competition for resources with the visual/manual task of active navigation. The data are interpretable with respect to resource theory such that if resource demand required to perform multiple tasks are drawn from the same limited resource pool (in this case, the visual modality), performance on all tasks may suffer (Wickens & Hollands, 2000). In the current experiment, the cost of interactivity (i.e., control precision) was observed when the cue was unreliable and the visual system was taxed not only with performing the bottom-up target detection task but also the target-distractor discriminability task when the cue was only partially reliable.

While interactivity was thereby influential in modulating the effects of trust, it also exhibited some direct effects. The discussion now turns to consider the direct effects of interactivity.

8.4 Interactivity

Egomotion is often cited as a critical cue for increasing the realism of a virtual environment (Rinalducci, 1996; Wickens & Baker, 1995). Additionally, allowing for interactivity may enhance this sense of realism by keeping the user consistently in the loop (Sheridan, 1996). Subjective data from the present experiment suggests that while the addition of active navigation did not explicitly result in an increased sense of realism within the simulation, it did serve to keep the user more involved or interested in such a way that active navigators felt marginally less workload (as revealed by mental demand, temporal demand, and effort ratings on the NASA-TLX scale) than passive viewers. In fact, observation of the subjects as they performed the experimental tasks revealed that active navigators viewed the simulation more as a game.

Although interactivity was not hypothesized to directly influence trust in the cueing information (except as mediated by realism, an effect that was not observed), it *was* expected to modulate attention allocation such that the workload imposed by the navigation task would make subjects more reliant on the cueing information, without necessarily increasing trust in that information. In fact, increasing workload has been found to lead to greater reliance (though not necessarily greater trust) in automation (McFadden et al., 1998; Parasuraman & Riley, 1997; Riley, 1994). However, given that subjects did not consciously notice the additional workload imposed by active navigation, it is not surprising that active navigation did not increase reliance on the cueing information in the present experiment. In fact, the opposite was found: passive viewers were more likely to be complacent or overreliant on automation. While this result is somewhat surprising, Wickens and Kessel (1980) suggested that the very act of participation serves as a motivator and as such, could benefit performance on other tasks performed concurrently.

The task of active navigation was expected to direct attention to aspects of the environment relevant to the active control task. If interactivity indeed moderated attention allocation and trust, then the data should show:

1) greater allocation of attention to the center of the display,

2) greater reliance on the database and its progeny (i.e., the cueing information), and

3) greater interest and allocation of attention to all tasks.

The data reveal that interactivity did not modulate attention allocation (as was suggested by Flach, 1990 and Ito & Matsunaga, 1990); rather head scanning data in Table 7.3.1 revealed that subjects allocated attention to similar areas in both passive and active conditions. Although it may be argued that one may simply move the eyes instead of moving the head and that a more precise means of measuring areas of interest should be used (e.g., eye tracking), the performance data confirm the finding revealed by the head movements. That is, there was no difference in target detection distance nor accuracy due to interactivity, nor any differential effects of interactivity on detecting the different salience targets, despite the fact that land mines were more in the center of the pathway and tanks and soldiers in the periphery.

The results are thus inconsistent with previous findings that active navigation should call attention to features in the environment closer to regions relevant to the control task (in this case, the land mines located along the path). Wallis and Bülthoff (1998) reported that drivers detected more obstacles when they were presented in the roadway than in the periphery. However, in their study, the roadway targets were more likely to represent hazards to be avoided in the navigation task than the more peripheral objects whereas in the current experiment, all targets were equally hazardous. Additionally, the target detection task our subjects performed was relatively difficult as the targets were not salient. Consequently, the use of top-down processing, in which task expectancies guide search in a way that attention would be allocated more to the task-relevant aspects of navigation, may have been inefficient due to the ambiguousness of the stimuli camouflaged in the terrain, and instead, subjects relied on more basic bottom-up processes.

Finally, performance of the local positioning task, which required locating checkpoints and verifying the validity of the terrain data through terrain association, may have modulated attention allocation, directing subjects' attention to all areas of the terrain equally, regardless of interactivity. This interpretation is consistent with previous studies, which have found that when subjects are instructed to visually explore the environment, regardless of interactivity, attention biases are eliminated (Shebilske et al., 1998; Wilson et al., 1997). However, the terrain association task was difficult and the data suggests that subjects in neither group (active or passive) performed it well.

8.5 Image Realism

Image realism in the simulation was implemented in two ways: (1) increasing the detail of the texture on the surface, and (2) increasing the number of polygons with which the environment was depicted. The data reveal that the manipulations were effective; subjective ratings indicate that the high detail scene was more involving and more compelling. In fact, subjects who had previously participated in training maneuvers at Ft. Irwin recognized various areas of the terrain depicted in the high detail simulation. However, neither performance data nor subjective ratings suggest that greater realism engendered greater trust. If, following the implications of Theunissen's (1998) hypothesis, image realism had indeed moderated the effects of cueing in a way attributable to biases in trust in the cueing information, trust in the simulation data, or attention allocation, the data would have revealed:

(1) greater trust in the database and all its progeny (i.e., the cueing information)

(2) changed salience of uncued targets

(3) a greater attentional cost or a greater tendency to fail to notice the missing landmarks when they occurred in the high detail scene.

Although the data do show a significant benefit of cueing, especially in the high detail terrain (Figure 7.2.1), were this effect mediated by trust or attention, the data should have also shown a greater attentional tunneling effect, i.e., a greater tendency to fail to notice the missing landmarks when they occurred in the high detail scene. Instead, the effects of scene detail did not appear anywhere else in the data, suggesting that the greater cueing benefits in the high detail scene were due simply to the fact that detection of the uncued targets were more difficult in the high detail scene *particularly when they were low in salience* (i.e., those targets had more room to benefit from cueing).

From the standpoint of performance and subjective data, high realism did not engender greater trust. Rather the differences in performance were attributable to subjects' setting of ß, in such a way that subjects were riskier in the low detail scene. More importantly though, the signal detection data suggests that realism did influence reliance on the cueing information. As shown in Figure 7.4.3.1, in the high detail scene, subjects were reluctant to report an object as a target when objects were *uncued*; however, when cueing information became available, subjects' responded by substantially shifting their response criterion so that they were generally riskier in their assessments. The degree of this beta shift was greater than that observed in the low detail scene, suggesting that subjects were in general *more reliant* on cueing information in the high detail scene than in the low detail scene. The findings replicate those of MacMillan et al. (1994), whose results suggested that operators would *overtrust* automation when the image quality was high, and lends support to Theunissen's hypothesis that trust in a highly realistic scene could be passed on to its progeny (the cueing information).

The issue of trust in the simulation data was measured implicitly by asking subjects to detect differences between the map and the terrain depicted, while performing a secondary task which required them to locate checkpoints in the terrain based on information on a hand-held paper map. Theunissen (1998) hypothesized that enhancing the visual fidelity of a simulation could result in inappropriate trust allocation, assuming that the information depicted was correct regardless of whether this trust was warranted or not, which could consequently result in unintended use of the information in the display. However, the present analysis revealed no difference in distance error in the detection of checkpoints resulting from scene fidelity, and more importantly, none of the subjects explicitly detected the missing mountain, the implicit measure of trust. Although two subjects failed to report the checkpoint next to the missing mountain, neither was sure as to the cause – e.g., the checkpoint may have been missed due to the fact that the mountain being searched for did not exist or that the tasks being performed were high in workload (e.g., both subjects were actively navigating in the high detail scene, in which target detection was more difficult). Thus, the data suggest that there was an *absence* of a tendency to exhibit this *overtrust* behavior in the high detail scene.

There are several explanations for the poor performance of all subjects on the implicit trust measure of noticing the missing mountain. First, it is possible that the nature of the task, one of detecting a missing object, was simply not salient enough. That is, the likelihood of detecting the presence of an unexpected object is greater than the likelihood of detecting the absence of an expected object (Wickens & Hollands, 2000). Secondly, when subjects were questioned about their strategy in performing the task, they indicated that they often did not perform terrain association but instead used the turns in the path closest to the checkpoint as their cue. No man-made features (e.g., buildings, cities, roads) were depicted in the simulation, and subjects felt that the presence of landmarks would have made discerning distance information from the terrain easier.

Third, it is also possible that subjects performed poorly on the checkpoint task because this task was not a high priority. When subjects were asked how they prioritized their different tasks, they cited target detection, navigating (if subjects were in the active condition), and then detecting the checkpoints. And in fact, some subjects were so busy with the target detection task that they forgot about the checkpoint task or commented that they were too busy to carefully evaluate differences between the terrain and the map.

While none of the subjects explicitly noticed the missing mountain, four subjects reported "errors" in which changes in the terrain, depicted on the map were not visible in the low detail terrain. For example, three subjects commented that the terrain information was not defined in enough detail as gradual ascents and descents were not apparent in the terrain or if the changes were visible, then they were too sharp, e.g., the mountains on the map seemed steeper in the simulation than they were depicted on the map. These errors could be attributable to the averaging of polygons, the algorithm used to generate the low detail terrain, and this averaging minimized critical differences in the terrain simulation relative to the map. Thus, the implicit trust measures suggest that there was an effect of realism, revealing that in developing a simulation for land navigation and terrain awareness, what is important may not be how terrain is textured, but rather the detail with which the terrain information is generated, i.e., the number of polygons used to generate the terrain data.

8.6 Model of Attention and Trust Biases

The model presented in Figure 8.1.1 was revised in certain respects given the results revealed by the data. Figure 8.6.1 is an effort to characterize how the four independent variables shown on the left (cueing, interactivity, reliability, and realism) are influenced by four key cognitive variables listed on the right which appear to have mediated the various performance effects that were observed, two of these accommodated in the original model – *visual attention* and how it was distributed through the environment (assessed using target detection distance and accuracy measures) and *trust* in the reliability of an information source which often determines how efficiently the automated information is used; and two incorporated into the revision – *resource demand* and its distribution to different tasks and *visibility*, the target's appearance and contrast in the environment.

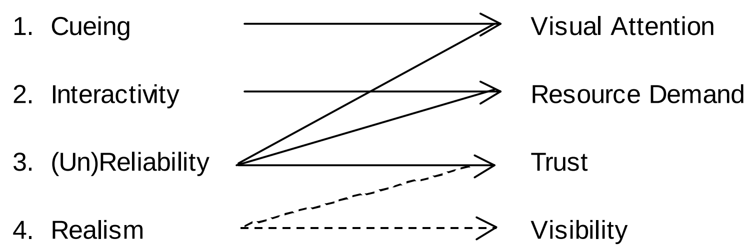


Figure 8.6.1. Cognitive factors mediated by the use of cueing, image realism, reliability, and interactivity in display design. The solid arrows describe those variables on the left which *directly* influence the cognitive factors on the right; the dotted arrows represent *indirect* influences.

In Figure 8.6.1, the solid arrows describe *direct* influences between the independent variables examined in the current study listed on the left and the cognitive factors on the right. The dotted arrows represent *indirect* relationships.

Cueing served to modulate *attention* allocation through the scene, aiding visual search by directing attention to targets in the environment. Its benefits to detection were observed in particular for those targets of low *visibility*, i.e., low salience (as shown in Figure 7.1.1(b)). More importantly though, the data reveals a cost attributable to cueing for the detection of the simultaneous uncued targets...the high priority nuclear device, which was never cued directly but was presented in conjunction with a cued or uncued soldier. In this case, cueing guided attention *away* from the high priority event, and the unexpected event was detected *less* when it was presented with a cued soldier than with an uncued soldier (as shown in Figure 7.1.1). This attentional tunneling behavior has been reported elsewhere (Merlo et al., 1999; Yeh et al., 1999) and is symptomatic of an attentional bias possibly resulting from cognitive laziness (Mosier et al., 1998).

The benefits of cueing were reduced when the cue was *unreliable* thus extending the findings of Posner and Snyder (1975) of a cost-benefit relationship to the presentation of cueing, to automated cueing in a noisy environment. In their study, benefits, calculated to be the reduction in detection time when a cue was presented versus no cue, increased at a faster rate than the cost, calculated as the difference in detection time with no cue versus an incorrect cue, when the cueing was highly (but not totally) reliable. In the current study, this knowledge of the cue's *unreliability* served to modulate *attention allocation* and distribution through the scene by increasing subjects' attentional breadth so that they were less reliant on the cueing information (a cost to detecting cued targets) and more likely to scan areas around that which was cued in order to detect high priority events (Figure 7.4.4.1), a result also reported by Merlo et al. (1999). That is, *trust* in the cueing information as operationally defined by the narrow attention focus was attenuated by unreliability.

However, the high reliability of the cue (even when it was not perfectly so) still induced a high level of *trust* (as hypothesized by Muir, 1987 and shown by Lee & Moray, 1994), which was unwarranted given the actual precision of the cueing information. The overtrust, exhibited by the greater false alarm rate – the incorrect detection of cued distractors – when non-targets were cued than when they were not, shows that subjects became somewhat complacent, attending more to the hypothesis suggested by the cue rather than to the underlying visual data (the target shape) itself, replicating the results of Conejo and Wickens (1997). These results are also consistent with a framework proposed by Wickens et al. (1999), which predicts that attention allocation strategies would be significantly influenced by the human's *trust* in the cueing information. While this did provide benefits (increased target detection distances), it was not failsafe (higher rate of false alarms and less attention to the area surrounding a cue – the location of the unexpected high priority targets – than to the cued area itself).

Additionally, *unreliability* in cueing influenced *resource demand*. Since subjects were unable to unquestioningly rely on the cue but needed to more closely examine the information highlighted by the cue, the ability to distribute visual attention effectively through the environment was taxed and attention was distributed less efficiently. The cost of this resource demand became apparent when interactivity was introduced. Subjects who actively navigated were less able to attend to the target detection and discriminability task and still maintain control precision when the cueing was only partially reliable. Thus, the data suggest a strong trend in a pattern consistent with other effects showing that active subjects exhibited less overtrust of unreliable automation (fewer false alarms when it was unreliable) but also exhibited more undertrust (i.e., fewer cueing benefits), and their greater reliance on the more difficult to process raw data imposed resource costs that interfered more with the resource demands of active navigation. However, the slight increase in resource demand imposed by active navigation was revealed only through performance data (the decrement in target detection distance when it was paired with unreliability) and not revealed through subjective workload measures, a dissociation reported previously by Yeh and Wickens (1988).

Finally, the data reveal a cost for increasing *realism*, which when implemented using the techniques employed here, served to increase the noise in the visual scene and reduce the *visibility* of targets camouflaged in the terrain if they were low in salience (Figure 7.2.1). However, signal detection data reveal no difference in sensitivity in subjects' ability to detect the high salient objects; rather, subjects responded to the change in image realism through a beta shift, becoming riskier in their responses in the low detail scene relative to the high detail one. That is, the higher rate of hits in the low detail scene was accompanied by a higher rate of false alarms than in the high detail scene.

Although subjects did not trust the terrain information in the simulation any more in the high detail scene than they did in the low detail scene, an interesting result was revealed in its interaction with cueing. If a shift (decrease) in β when cueing is imposed can be operationally defined as trust in the cueing information, then the data suggests that as image quality is increased, subjects trust the cue more. As shown in Figure 7.4.3.1, the β shift as the cue becomes more informative and more reliable is greater in the high detail terrain than in the low. Thus, while the data do not confirm realism's direct link to trust, they do does suggest that the progeny of the realistic rendering – the attentional guidance symbology – may inherit properties implicit in the generation of a high fidelity display so that operators trust the guidance information more than when it appears in a lower fidelity scene.

9. CONCLUSIONS

9.1 Theoretical Implications

The current research informs theories of attention deployment and the cost-benefit analysis of this information. The presentation of cueing served as an effective method for deploying attention and the benefits were more strongly realized when it was 100% reliable (as predicted by Posner & Snyder, 1975). However, cueing decreased subjects' attentional breadth, guiding attention away from critical targets present in the environment, and resulted in tunneling on the area highlighted by the automation. More interestingly, this attention allocation behavior was moderated by cue reliability and *trust* in the cueing information (as suggested by the results of Merlo et al. (1999)). Knowledge that the cue was only partially reliable *reduced* tunneling on the regions in space highlighted by the cue and increased subjects' attentional breadth in such a way that they were more likely to detect the unexpected, high priority event, *regardless of whether or not it was presented with a cued or uncued object*. Unfortunately, partially reliable cueing had other costs, e.g., increased false alarm rate due to overtrust in the cueing information and reduced precision in control as a consequence of insufficient resource availability. Indeed, the loss of reliability appeared to eliminate any of the benefits of cueing insofar as sensitivity was concerned.

Thus, the examination of the effectiveness of partially reliable cueing adds to theories of automation trust set forth by Muir (1987) and more specifically in examining interaction with automation, theories set forth by Lee and Moray (1992, 1994) and Parasuraman and Riley (1997). The significant increase in false alarms when the distractors were cued reveal that trust in cueing information was not calibrated optimally, even when subjects were informed as to the unreliability of the automation. This reliance was reduced by increased interactivity but increased through enhanced scene detail. The benefits of interactivity for reducing automation-induced complacency is not new (Parasuraman & Riley, 1997; Parasuraman et al., 1996), but it is interesting that interactivity and realism, both features believed to enhance the sense of involvement within a simulation (Hendrix & Barfield, 1995; Rinalducci, 1996), would have such opposite effects. In fact, signal detection analysis suggests that realism's link to trust (especially overtrust) as hypothesized by Theunissen (1998) may be passed on to the progeny of the realistic rendering (i.e., the cue upon which it is overlaid) suggesting that realism may have implicit (and potentially costly) influences on trust and attention deployment throughout a visual scene; this is an effect that needs to be studied further.

9.2 Design Implications

In addition to providing an improved understanding of the attentional mechanisms and trust biases underlying use of automation tools, some design implications can be drawn from this research. First, careful design and testing of features intended to enhance reality is required, as evidenced by the strong interaction between visibility/target saliency and visual noise, shown in Figure 7.2.1. The result clearly stresses that enhancing reality is not a prerequisite for good performance and that before implementation of these display features, the task-dependent constraints need to be considered. For example, when critical information about the terrain needs to be conveyed, generating the scene with a high number of polygons is more critical than increasing the detail with which the scene is textured. In the current study, subjects reported false errors in the terrain database due to the averaging of polygons used to generate the low detail scene which caused some areas of the terrain to appear more mountainous (or less) than that depicted on the paper map. On the other hand, if the task is simply one of exploration or influencing one's sense of "being there", enhancing the texture of the visual scene certainly makes the simulation more compelling and more involving.

Interactivity served to immerse the user in the environment and task to such a degree that active navigators subjectively experienced less mental effort and frustration than passive viewers. This was achieved by the effort we gave to employing an easy-to-use intuitive navigational interface. It must be noted, however, that the navigation task was implemented in such a way to aid control precision, e.g., a tunnel pathway which has been shown to improve precision relative to less direct means of flight path presentation (Ververs & Wickens, 1999), and hence did not impose a significant amount of workload. However, costs to task performance should be expected when the difficulty of this control task is increased (as suggested by the results of Williams et al., 1996, Experiment 1) or when task load is increased, e.g., when attention was guided by potentially unreliable cueing information in the current experiment.

With respect to the presentation of cueing information, the data certainly suggest that trust in the automated cueing information influenced attention allocation strategies, and the cost to perfectly reliable cueing is observed in the detection of the nuclear devices. Partially reliable cueing was able to compensate for some of these effects but had other costs, e.g., increased false alarm rate due to overtrust in the cueing information and reduced precision in control as a consequence of insufficient resource availability. It is unreasonable to assume that 100% reliable cueing will be possible in operational environments, i.e., there will be errors in the system, and unfortunately, the data suggests that if such an error is present in the system, subjects will not detect it. One option would be to present the cue in such a way that it guides attention to a broader area of the display space, rather than directly to the target (Merlo et al., 1999); such symbology could widen subjects' attentional breadth.

Thus, the results of the study emphasize the need for designers of such cueing systems to more carefully evaluate operator reliance on automation. The ability to provide highly precise cueing information – as was implemented in the current study – is subject to factors such as time delay, incorrect reports from intelligence analysts or sensors, and errors in head-tracking from automation aids, and this imprecision may make the cueing data less informative. Consequently, the designer needs to examine in particular the occurrence of false alarms – and the cost of an incorrect detection, given that the presentation of cueing will increase the likelihood of these errors. This is an issue that needs to be carefully considered in the implementation of automated cueing.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Polly Baker, Dave Irwin, Art Kramer, and Nadine Sarter for their suggestions in the development of this experimental study and in the manuscript. We are especially grateful to Polly Baker, Albert Khakshour, and Rob Stein for allowing us to use the software they developed for terrain battlefield visualization and we are very appreciative of their efforts in customizing the simulation for the purposes of the current study. Special thanks are also due to the talented programmer, Ron Carbonari, for his time and patience in the development of the data collection software, and David Brandenburg for his invaluable help in running the study. Finally, this experiment would not have been possible without all the Army, Army National Guard, and Marine personnel who participated in the study. We would especially like to thank Lieutenant Colonel E.A. Charlie Hart and Sergeant Kurt Sharp from the University of Illinois Reserve Officers Training Corps and First Sergeant Michael McMullen and Captain Ed Lockwood from the National Guard Lincoln’s Challenge program in Rantoul, Illinois for their assistance in recruiting participants.

This material is based upon work supported by the U.S. Army Research Laboratory under Award No. DAAL 01-96-2-0003. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Army Research Laboratory.

REFERENCES

Appleyard, D. (1970). Styles and methods of structuring a city. *Environment and Behavior, 2*, 100-118.

Arthur, E. J., Hancock, P. A., & Chrysler, S. T. (1993). Spatial orientation in real and virtual worlds. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 328-332). Santa Monica, CA: Human Factors & Ergonomics Society.

Arthur, E. J., Hancock, P. A., & Chrysler, S. T. (1997). The perception of spatial layout in real and virtual worlds. *Ergonomics, 40*(1), 69-77.

Attneave, F. (1964). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183-193.

Ball, K. & Sekular, R. (1981). Cues reduce direction uncertainty and enhance motion detection. *Perception and Psychophysics, 30*(2), 119-128.

Balla, J. (1980). Logical thinking and the diagnostic process. *Methodology and Information in Medicine, 19,* 88-92.

Balla, J. (1982). The use of critical cues and prior probability in concept identification. *Methodology and Information in Medicine, 21*, 9-14.

Barfield, W., Rosenberg, C., & Furness, T. A. (1995). Situation awareness as a function of frame of reference, computer graphics eyepoint elevation, and geometric field of view. *International Journal of Aviation Psychology, 5*(3), 233-256.

Barfield, W., Rosenberg, C., & Kraft, C. (1990). Relationship between scene complexity and perceptual performance for computer graphics simulations. *Displays*, 179-185.

Barfield, W., Rosenberg, C., & Lotens, W. A. (1995). Augmented-reality displays. In W. Barfield & T. A. Furness (Eds). *Virtual environments and advanced interface design*. New York: Oxford University Press.

Becklen, R. & Cervone, D. (1984). Selective looking and the noticing of unexpected events. *Memory and Cognition, 11*(6), 601-608.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *Journal of the Acoustical Society of America, 88*(1), 149-158.

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. New Jersey: Lawrence Erlbaum.

Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence, 6*(1), 73-86.

Bossi, L. L., Ward, N. J., Parkes, A. M., & Howarth, P. A. (1997). The effect of vision enhancement systems on driver peripheral visual performance. In Y. I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 239-260). New Jersey: Lawrence Erlbaum Associates.

Broadbent, D. E. (1965). Information processing in the nervous system. *Science, 3695*, 457-462.

Brooks, F. P. (1988). Grasping reality through illusion: Interactive graphics serving science. *Proceedings of CHI '88* (pp. 1-11). New York: ACM.

Caro, P. W. (1988). Flight training and simulation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 229-261). San Diego, CA: Academic Press.

Carroll, J. M. & Olson, J. R. (1988). Mental models in human-computer interaction. In M. Helander (Ed.), *Handbook of human-computer interaction*. North-Holland: Elsevier Science Publishers.

Casey, S. M. (1993). *Set phasers on stun and other true tales of design, technology, and human error*. Santa Barbara, CA: Aegean.

Clawson, D. M., Miller, M. S., Knott, B. A., & Sebrechts, M. M. (1998). Navigational training in virtual and real buildings. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1427-1431). Santa Monica, CA: Human Factors & Ergonomics Society.

Conejo, R., & Wickens, C. D. (1997). *The effects of highlighting validity and feature type on air-to-ground target acquisition performance* (Technical Report ARL-97-11/NAWC-ONR-97-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Cowan, N. (1988). Evolving conceptions of memory storage, electronic storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*(2), 163-191.

Darken, R. P. & Sibert, J. L. (1996). Navigating large virtual spaces. *International Journal of Human-Computer Interaction, 8*, 49-71.

Davison, H. & Wickens, C. D. (1999). *Rotorcraft hazard cueing: The effects on attention and trust* (Technical Report ARL-99-5/NASA-99-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Devlin, A. S. & Bernstein, J. (1997). Interactive way-finding: Map style and effectiveness. *Journal of Environmental Psychology, 17,* 99-110.

Dopping-Hepenstal, L. L. (1981). Head-up displays: The integrity of flight information. *IEE Proceedings- F Communications Radar and Signal Processing*, 128F, 7, 440-442.

Drascic, D. & Milgram, P. (1996). Perceptual issues in augmented reality. In M. T. Bolas, S. S. Fisher, & J. O. Merritt (Eds.), *Proceedings of the International Society for Optical Engineers (SPIE): Stereoscopic Displays and Virtual Reality Systems III* (pp. 123-134). 2653. Bellingham, WA: SPIE.

Dwyer, F. M. (1967). Adapting visual illustrations for effective learning. *Harvard Educational Review, 37*, 250-263.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

Egeth, H. E. & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269-297.

Ellis, S. R., McGreevy, M. W., & Hitchcock, R. J. (1987). Perspective traffic display format and airline pilot traffic avoidance. *Human Factors, 29*(4), 371-382.

Ellis, S. R., Tharp, G. K., Grunwald, A. J., & Smith, S. (1991). Exocentric judgments in real environments and stereoscopic displays. *Proceedings of the 35th Annual Meeting of the Human Factors Society* (pp. 1442-1446). Santa Monica, CA: Human Factors Society.

Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394.

Entin, E. B. (1998). The effects of decision aid availability and accuracy on performance and confidence. *Proceedings Applied Behavioral Sciences Symposium*. Colorado Springs, CO: U.S. Air Force Academy.

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics, 16*, 143-149.

Evans, G. W. & Pedzek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory, 6*(1), 13-24.

Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors & Ergonomics Society.

Fisher, D. L., Coury, B. G., Tengs, T. O., & Duffy, S. A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors, 31*(2), 167-182.

Fisher, D. L. & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors, 31*, 17-30.

Flach, J. M. (1990). Control with an eye for perception: Precursors to an active psychophysics. *Ecological Psychology, 2*(2), 83-111.

Foreman, N., Foreman, D., Cummings, A., & Owens, S. (1990). Locomotion, active choice, and spatial memory in children. *Journal of General Psychology, 117*, 215-232.

Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E. M., Begault, D. R., & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system: Problem, design philosophy, and description of an integrated display suite for low-visibility airport surface operations. *Proceedings of the SAE/AIAA World Aviation Congress, Paper 965551*. Washington, DC: American Institute of Aeronautics and Astronautics.

Foyle, D. C., McCann, R. S., & Shelden, S. G. (1995). Attentional issues with superimposed symbology: Formats for scene-linked displays. In R. S. Jensen (Ed.), *Proceedings of the Eighth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Gale, N., Golledge, R. G., Pellegrino, J. W., & Doherty, S. (1990). The acquisition and integration of route knowledge in an unfamiliar neighborhood. *Journal of Environmental Psychology, 10*, 3-25.

Gempler, K. S. & Wickens, C. D. (1998). *Display of predictor reliability on a cockpit display of traffic information* (Technical Report ARL-98-6/ROCKWELL-98-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology, 49*, 182-194.

Gibson, J. J. (1962). Observations on active touch. *Psychological Review, 69*(6), 477-491.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gossweiler, R. (1994). A system for application-independent time-critical rendering. *ACM Proceedings of SIGCHI '94 Conference on Human Factors in Computing* (p. 261). New York: ACM.

Grieg, I., Shaw, G., & Lumsden, B. (1997). Augmented vision systems for the all weather autonomous landing guidance of fixed wing aircraft: An investigation of the key parameters by piloted flight simulation. *Proceedings of the 10th CEAS Conference*. Amsterdam, The Netherlands.

Hancock, P. A., Hendrix, C., & Arthur, E. (1997). Spatial mental representation in virtual environments. *Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1143-1147). Santa Monica, CA: Human Factors & Ergonomics Society.

Held, R. & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology, 56*(5), 872-876.

Hendrix, C. & Barfield, W. (1995). Presence within virtual environments as a function of visual and auditory cues. *Proceedings of the Virtual Reality Annual International Symposium* (pp. 74-82). Los Alamitos, CA: IEEE.

Henry, D. & Furness, T. (1993). Spatial perception in virtual environments: Evaluating an architectural application. *1993 IEEE Virtual Reality International Symposium* (pp. 33-40). Los Alamitos, CA: IEEE.

Hübner, R. (1996). The efficiency of different cue types for reducing spatial-frequency uncertainty. *Vision Research, 36*(3), 401-408.

Hughes, D. (1998, September 7). Avionics on verge of providing full weather, terrain picture. *Aviation Week and Space Technology*, 142-147.

Ito, H. & Matsunaga, K. (1990). Relative distance perception through expanding and contracting motion and the role of propriospecific information in walking. *Ecological Psychology, 2*(2), 113-130.

Jaeger, B. K. (1998). The effects of training and visual detail on accuracy of movement production in virtual and real-world environments. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1486-1490). Santa Monica, CA: Human Factors & Ergonomics Society.

Johnson, W. W. & Schroeder, J. A. (1995). Visual-motion cueing in the control of altitude. *IEEE*, 2676-2681.

Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology, 34*(2), 103-112.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

Kaber, D. M., Onal, E., & Endsley, M. R. (1998). Level of automation effects on telerobot performance and human operator situation awareness and subjective workload. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 165-170).

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors, 39*(2), 164-176.

Kaplan, R., Kaplan, S., & Deardorff, H. L. (1974). The perception and evaluation of a simulated environment. *Man-Environment Systems, 4*, 191-192.

Kerr, S. T. (1990). Wayfinding in an electronic database: The relative importance of navigational cues vs. mental models. *Information Processing and Management, 26*(4), 511-523.

Kleiss, J. A. (1995). Visual scene properties relevant for simulating low-altitude flight: A multidimensional approach. *Human Factors, 37*(4), 711-754.

Kleiss, J. A., Curry, D. G., & Hubbard, D. C. (1988). Effect of three-dimensional object type and density in simulated low-level flight. *Proceedings of the 32nd Annual Meeting of the Human Factors Society* (pp. 1299-1303). Santa Monica, CA: Human Factors Society.

Kleiss, J. A. & Hubbard, D. C. (1993). Effects of three types of flight simulator visual scene detail on detection of altitude change. *Human Factors*, *35*(4), 653-671.

Kramer, A. F. & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics*, *50*, 267-284.

Larish, I. & Wickens, C. D. (1991). *Divided attention with superimposed and separated imagery: Implications for head-up displays* (Technical Report ARL-91-4/NASA HUD-91-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Larish, J. F., & Andersen, G. J. (1993). Active control in interrupted dynamic spatial orientation: The detection of orientation change. *Perception and Psychophysics*, *37*(4), 533-545.

Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270.

Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*, 153-184.

Levine, M., Jankovic, I. N., & Palij, J. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General, 111*(2), 157-175.

Levy, J., Foyle, D. C., & McCann, R. S. (1998). Attentional allocation and head-up displays. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors & Ergonomics Society.

Lintern, G. & Garrison, W. V. (1992). Transfer effects of scene content and crosswind in landing instruction. *International Journal of Aviation Psychology, 2*(3), 225-244.

Lintern G. & Koonce, J. M. (1991). Display magnification for simulated landing approaches. *International Journal of Aviation Psychology, 1*(1), 59-72.

Lintern G. & Koonce, J. M. (1992). Visual augmentation and scene detail effects in flight training. *International Journal of Aviation Psychology, 2*(4), 281-301.

Lintern, G., Sheppard, D. L., Parker, D. L., Yates, K. E., & Nolan, M. D. (1989). Simulator design and instructional features for air-to-ground attack: A transfer study. *Human Factors, 31*(1), 87-99.

Lintern, G., Taylor, H. L., Koonce, J. M., Kaiser, R. H., & Morrison, G. A. (1997). Transfer and quasi-transfer effects of scene detail and visual augmentation in landing training. *International Journal of Aviation Psychology, 7*(2), 149-169.

Lintern, G., Thomley-Yates, K. E., Nelson, B. E., & Roscoe, S. N. (1987). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack. *Human Factors, 29*(1), 45-59.

Liu, Y., Fuld, R., & Wickens, C. D. (1995). Monitoring behavior in manual and automated scheduling systems. *International Journal of Man-Machine Systems, 39*, 1015-1029.

MacMillan, J., Entin, E. B., & Serfaty, D. (1994). Operator reliance on automated support for target recognition. *Proceedings of the 38th Annual Meeting of the Human Factors and*

Ergonomics Society (pp. 1285-1289). Santa Monica, CA: Human Factors & Ergonomics Society.

Mann, T. L. (1987). Autonomous landing guidance concept – The effects of video and symbology dynamics on pilot performance. *Proceedings of the Sixth Aerospace Behavioral Engineering Technology Conference* (pp. 131-138). SAE Technical Paper Series 872390.

Martin-Emerson, R. & Wickens, C. D. (1992). The vertical visual field and implications for the head-up display. *Proceedings of the Human Factors Society*. Santa Monica, CA: Human Factors & Ergonomics Society.

Martin-Emerson, R. & Wickens, C. D. (1997). Superimposition, symbology, visual attention, and the head-up display. *Human Factors, 39*(4), 581-601.

McCann, R. S., & Foyle, D. C. (1995). Scene-linked symbology to improve situation awareness. *1995 NATO - AGARD Symposium on Situation Awareness* (pp. 16-1 - 16-11). Neuilly Sur Seine, France, NATO Advisory Group for Aerospace Research and Development (NTIS No. AGARDCP 575).

McCann, R. S., Foyle, D. C., & Johnston, J. C. (1992). Attentional limitations with head-up displays. *SID International Symposium Digest of Technical Papers,* XXII.

McCormick, E. P., Wickens, C. D., Banks, R., & Yeh, M. (1998). Frame of reference effects on visualization subtasks. *Human Factors, 40*(3), 443-451.

McFadden, S. M., Giesbrecht, B. L., & Gula, C. A. (1998). Use of an automatic tracker as a function of its reliability. *Ergonomics, 41*(4), 512-536.

Merlo, J. L., Wickens, C. D., & Yeh, M. (1999). *Effect of reliability on cue effectiveness and display signaling* (Technical Report ARL-99-4/ARMY-FED-LAB-99-3). Savoy, IL: University of Illinois, Aviation Research Laboratory

Milgram, P. & Colquhoun, H. (1999). A taxonomy of real and virtual world display integration. In Y. Ohta & H. Tamura (Eds.), *Mixed reality – Merging real and virtual worlds*. Ohmsha (Tokyo) and Springer-Verlag (Berlin).

Mohindra, N. K, Spencer, E., & Lambert, A. (1986). Automaticity and the capture of attention by a peripheral display change. *Defense Technical Information Center Technical Report*. Defense Information Systems Agency.

Montgomery, D. A. & Sorkin, R. D. (1996). Observer sensitivity to element reliability in a multielement visual display. *Human Factors, 38*(3), 484-494.

Mosier, K. L., Palmer, E. A., & Degani, A. (1992). Electronic checklists: Implications for decision making. *Proceedings of the 36th Annual Meeting of the Human Factors Society* (pp. 7-11). Santa Monica, CA: Human Factors Society.

Mosier, K. L. & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum Associates.

Mosier, K., Skitka, L., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high technology cockpits. *International Journal of Aviation Psychology, 8,* 47-63.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*, 527-539.

Muir, B. M. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*(3), 429-460.

Müller, H. J. & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 315-330.

Nagy, A. L, & Sanchez, R. R. (1992). Chromaticity and Luminance as Coding Dimensions in Visual Search. *Human Factors, 34*(5), 601-614.

National Research Council. (1997). *Tactical display for soldiers: Human factors considerations*. Washington DC: National Academy Press.

Neisser, U. & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*, 480-494.

Ockerman, J. J. & Pritchett, A. R. (1998). Preliminary study of wearable computers for aircraft inspection. In G. Boy, C. Graeber, & J. M. Robert (Eds.), *HCI-Aero '98: International Conference on Human-Computer Interaction in Aeronautics*.

Olmos, O., Wickens, C. D., & Chudy, A. (1997). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts, and the application of cognitive engineering. In R. S. Jensen (Ed.), *Proceedings of the 9th International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Padmos, P. & Milders, M. V. (1992). Quality criteria for simulator images: A literature review. *Human Factors, 34*(6), 727-748.

Parasuraman, R., Molloy, R., Moulou, M., & Hilburn, B. (1996). Monitoring of Automated Systems. In R. Parasuraman & M. Moulou (Eds.), *Automation and human performance* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*(1), 1-23.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Pausch, R., Burnette, T., Brockway, D., & Weiblen, M. E. (1995). Navigation and locomotion in virtual worlds via flight into hand-held miniatures. *Computer Graphics: ACM SIGGRAPH '95 Conference Proceedings*.

Péruch, P. & Gaunet, F. (1998). Virtual environments as a promising tool for investigating human spatial cognition. *Current Psychology of Cognition, 17*(4-5), 881-899.

Péruch, P., Vercher, J. L., & Gauthier, G. M. (1995). Acquisition of spatial knowledge through visual exploration of simulated environments. *Ecological Psychology, 7*(1), 1-20.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I. & Snyder, C. R. R. (1974). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic, *Attention and performance V*. New York: Academic Press.

Posner, M. I. & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160-174.

Presson, C. C., DeLange, N., & Hazelrigg, M. D. (1989). Orientation specificity in spatial memory: What makes a path different from a map of the path. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 887-897.

Presson, C. C. & Hazelrigg, M. D. (1984). Building spatial representations through primary and secondary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 716-722.

Proctor, R. W. & van Zandt, T. (1994). *Human Factors in Simple and Complex Systems*. Boston, MA: Allyn and Bacon.

Reardon, K. A. (1988). The effects of nested texture on a landing-judgment task. *Proceedings of the 32nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 10-14). Santa Monica, CA: Human Factors and Ergonomics Society.

Regal, D. M. & Whittington, D. H. (1994). Synthetic vision display evaluation studies. *NASA Contractor Report 194963.*

Riley, V. (1994). *Human Use of Automation*. Doctoral dissertation: University of Minnesota, Minneapolis, MN.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications. Human factors in transportation* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rinalducci, E. J. (1996). Characteristics of visual fidelity in the virtual environment. *Presence, 5*(3), 330-345.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia, 25*, 31-40.

Rolland, J. P., Biocca, F. A., Barlow, T., & Kancerla, A. (1995). Quantification of adaptation to virtual-eye location in see-through head-mounted displays. *1995 IEEE Virtual Reality International Symposium* (pp. 56-66). Los Alamitos, CA: IEEE.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in “desk-top” virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology, 3*(2), 143-159.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *International Journal of Aviation Psychology, 8*(3), 197-208.

Sampson, J. B. (1993). Cognitive performance by individuals using a head-mounted display while walking. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors & Ergonomics Society.

Sarter, N. B. & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots’ model and awareness of the flight management system. *The International Journal of Aviation Psychology, 4*(1), 1-28.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed.). New York: Wiley.

Satalich, G. A. (1995). Navigation and wayfinding in virtual reality: Finding the proper tools and cues to enhance navigational awareness. Unpublished masters thesis: University of Washington.

Schum, D. A. (1975). The weighting of testimony in judicial proceedings from sources having reduced credibility. *Human Factors, 17*(2), 172-182.

Schwartz, A. S., Perey, A. J., & Azulay, A. (1979). Further analysis of active and passive touch in pattern discrimination. *Bulletin of the Psychonomic Society, 6*(1), 7-9.

Shebilske, W. L., Jordan, J. A., Goettl, B. P., & Paulus, L. E. (1998). Observation versus hands-on practice of complex skills in dyadic, triadic, and tetradic training-teams. *Human Factors, 40*(4), 525-540.

Sheridan, T. B. (1996). Further musings on the psychophysics of presence. *Presence, 5*(2), 241-246.

Singer, M. J. & Witmer, B. G. (1996). *Presence measures for virtual environments: Background & development* (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Singh, I. L., Molloy, R. & Parasuraman, R. (1993). Individual differences in monitoring failures of automation. *The Journal of General Psychology, 120*(3), 357-373.

Slater, M. & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence*, *6*(6), 603-616.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84*(3), 1107-1108.

Stappers, P. J. (1989). Forms can be recognized from dynamic occlusion alone. *Perceptual and Motor Skills*, *68*, 243-251.

Stoakley, R., Conway, M. J., & Pausch, R. (1995). Virtual reality on a WIM: Interactive worlds in miniature. *Chi '95 Mosaic of Creativity* (pp. 265-272). New York: ACM.

Surdick, R. T., Davis, E. T., King, R. A., & Hodges, L. F. (1997). The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence, 6*(5), 513-531.

Suter, M. & Nüesch, D. (1995). Automated generation of visual simulation databases using remote sensing and GIS. *Proceedings Visualization '95* (pp. 86-93). Los Alamitos, CA: IEE.

Swennsen, R. G., Hessel, S. J., & Herman, P. G. (1977). Omissions in radiology: Faulty search of stringent reporting criteria? *Radiology, 123*, 563-567.

Theunissen, E. (1998). Spatial terrain displays: Promises and potential pitfalls. Presented at the *17th Digital Avionics Systems Conference,* October 31-November 6. Seattle, WA.

Thorndyke, P. W. & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology, 14*, 560-589.

Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology*, *40A*(2), 201-237.

Umiltá, C., Riggio, L., Dascola, I., & Rizzolatti, G., (1991). Differential effects of central and peripheral cues on the reorienting of spatial attention. *European Journal of Cognitive Psychology, 3*(2), 247-267.o

Ververs, P. M. & Wickens, C. D. (1998a). *Conformal flight path symbology for head-up displays: Defining the distribution of visual attention in three-dimensional space* (Technical Report ARL-98-5/NASA-98-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Ververs, P. M. & Wickens, C. D. (1998b). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology*, *8*(4), 377-403.

Vincow, M. (1997). Frame of reference and navigation through document visualizations. Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.

Vincow, M. & Wickens, C. D. (1998). Frame of reference and navigation through document visualizations: Flying through information space. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 511-515). Santa Monica, CA: Human Factors and Ergonomics Society.

Von Wright, J. M. (1957). A note on the role of “guidance” in learning. *British Journal of Psychology, 48*, 133-137.

Walk, R. D., Shepherd, J. D., & Miller, D. R. (1988). Attention and the depth perception of kittens. *Bulletin of the Psychonomic Society*, *26*, 248-251.

Wallis, G. & Bülthoff, H. H. (1998). Scene analysis whilst traveling a virtual roadway.

Watson, B., Walker, N., Hodges, L. F., & Worder, A. (1997). An evaluation of level of detail degradation in head-mounted display peripheries. *Presence, 6*(6), 630-637.

Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., & Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence, 5*(3), 263-273.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: Harper-Collins Publishers.

Wickens, C. D. (1997). Attentional issues in head-up displays. In *Engineering psychology and cognitive ergonomics: Integration of theory and application*. London: Avebury Technical Pub. Co.

Wickens, C. D. (1999). Frame of reference for navigation. In D. Gopher & A. Koriat (Eds.), *Attention and performance*, Vol. 17. Orlando, FL: Academic Press.

Wickens, C. D. & Baker, P. (1995). Cognitive issues in virtual reality. In W. Barfield & T. A. Furness (Eds), *Virtual environments and advanced interface design* (pp. 514-541). New York: Oxford University Press.

Wickens, C. D., Gempler, K., & Morphew, M. E. (in press, 2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal.*

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D. & Kessel, C. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 24-34.

Wickens, C. D. & Kessel, C. (1980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology: Human Perception and Performance, 6*(3), 564-577.

Wickens, C. D. & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied*, *1*(3), 179-193.

Wickens, C. D. & Prevett, T. T. (1995). Exploring the dimensions of egocentricity in aircraft navigation displays. *Journal of Experimental Psychology: Applied, 1*(2), 110-135.

Wickens, C. D., Pringle, H. L., & Merlo, J. (1999). *Integration of information sources of varying weights: The effect of display features and attention cueing* (Technical Report ARL-99-2/FEDLAB-99-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., Thomas, L., Merlo, J., & Hah, S. (1999). Immersion and battlefield visualization: Does it influence cognitive tunneling? *Proceedings of the 3rd Annual FedLab Symposium: Advanced Displays and interactive Displays Consortium* (pp. 111-115). Adelphi, MD: ARL Federated Laboratory.

Wiener, E. (1980). Flight-deck automation: promises and problems. *Ergonomics, 23*(10), 996-1011.

Williams, H. P., Hutchinson, S., & Wickens, C. D. (1996). A comparison of methods for promoting geographic knowledge in simulated aircraft navigation. *Human Factors, 38*(1), 50-64.

Wilson, P. N., Foreman, N., Gillett, R., & Stanton, D. (1997). Active versus passive processing of spatial information in a computer-simulated environment. *Ecological Psychology, 9*(3), 207-222.

Wilson, P. N., Foreman, N., & Tlauka, M. (1997). Transfer of spatial information from a virtual to a real environment. *Human Factors*, *39*(4), 526-531.

Witmer, R. G., Bailey, J. H., Knerr, B. W., & Parsons, K. C. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human Computer Studies*, *45*, 413-428.

Witmer, R. G. & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225-240.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295-340.

Yeh, M. & Wickens, C. D. (1998). *Visual search and target cueing: A comparison of head-mounted versus hand-held displays on the allocation of visual attention* (Technical Report ARL-98-2/ARMY-FED-LAB-98-2). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Yeh, M., Wickens, C. D., & Seagull, F. J. (1998). *Effects of frame of reference and viewing condition on attentional issues with helmet-mounted displays* (Technical Report). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target cueing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors, 41*(4), 524-542.

Yeh, Y-Y. & Wickens, C. D. (1988). The dissociation of subjective measures of mental workload and performance. *Human Factors, 30*, 111-120.

APPENDIX 1: STUDIES OF TRUST AND ATTENTIONAL BIASES

A1.1 Effects of Automation

Study: Conejo and Wickens (1997)
Domain: Air
Task: Identify objects as target or non-target given map and cueing information while performing an aerial bombing run.
Bias: Attention, Trust
Reliability: 70% or 40%
Reliability Display: None
Cueing: Target highlighted in red or target highlighted with lead-in blinking
Display source: Evans and Sutherland display
Independent display variables: Highlighting vs. no highlighting
Other Variables / Manipulations: Lead-in feature type (natural vs. cultural)
Target feature type (natural vs. cultural)

Description of Results:
The results showed that the presentation of valid highlighting led to increased confidence in pilots' assessment of whether or not to shoot a target but there was no increase in accuracy. The presentation of invalid highlighting led pilots down a garden path, such that they allocated less attention to determining the presence/absence of other items in the environment even when the automated cue was incorrect.

Study: Davison and Wickens (1999)
Domain: Air
Task: Simulated exploration mission requiring hazard (target) avoidance
Bias: Attention, Trust
Reliability: 100% or 70%
Reliability Display: None
Cueing: Target presented on electronic map and highlighted in the visual scene
Display source: Evans and Sutherland display
Independent display variables: Highlighting vs. no highlighting
Other Variables / Manipulations: Type of highlighting (flashing vs. intensification)

Description of Results:
The presentation of valid cueing resulted in greater hazard awareness for those objects which were cued, but directed attention away (measured in terms of response time) from unexpected events, not displayed on the map but presented in conjunction with highlighted hazards. Unreliability (i.e., a failure to depict a hazard) delayed hazard detection but did not lead to any near misses or collisions. A consequence of this unreliability was more vigilant monitoring of the environment and earlier initiation of maneuvers for hazard avoidance.

Study: Gempler and Wickens (1998)
Domain: Air
Task: Avoid traffic conflicts
Bias: Attention, Trust
Reliability: 100% self; 83% intruder
Reliability Display: Predictor information that displayed a 95% confidence interval of future position versus a single line predictor
Cueing: Length of the predictor line decreased as time to predicted conflict increased. When actual conflict was present, intruder highlighted in yellow.
Display source: workstation display
Independent display variables:
Other Variables / Manipulations: Traffic geometry (vertical and longitudinal)

Description of Results:
The presentation of predictor information was so compelling that subjects attended to what the aircraft would do in the future rather than what it was currently doing. The presentation of the wedge did not influence pilots' trust calibration or perceived reliability, potentially due to increased clutter or simply pilot strategies in using the display.

Study:	Kantowitz, Hanowski, and Kantowitz (1997)	**Domain:**	Driving
Task:	Navigation through familiar and unfamiliar environments	**Bias:**	Trust
Reliability:	100%, 71%, or 43%		
Reliability Display:	None		
Cueing:	None		
Display source:	Driving simulator consisting of two personal computers and two video displays		
Independent display variables:	Trust		
Other Variables / Manipulations:	Familiarity of the area		

Description of Results:
Subjective opinion and trust in the automated guidance system decreased as reliability decreased. Trust was regained as more accurate information was presented but this gain in trust was less likely at low reliability levels. The presentation of inaccurate information was more detrimental in terms of trust in the system when the driver was familiar with the environment and consequently, drivers in these settings relied on the automation less since they felt more confident in their knowledge about the area.

Study:	Liu, Fuld, and Wickens (1993)	**Domain:**	Generic
Task:	Assign “customers” to the shortest of three service lines	**Bias:**	Trust
Reliability:	100%		
Reliability Display:	None		
Cueing:	None		
Display source:	workstation monitor		
Independent display variables:	Active control versus passive monitoring for system errors		
Other Variables / Manipulations:			

Description of Results:
Subjects detected more errors when they were controlling the system actively but the results showed that subjects were overconfident in their ability to monitor the system relative to allowing the automation to perform the task, when in fact the two were equal.

Study:	McFadden et al. (1998)	**Domain:**	Generic
Task:	Monitor a set of moving targets and update their position based on new data regarding positions of current targets as well as the presence of new targets	**Bias:**	Attention
Reliability:	defined in terms of the size of the area around the target in which a signal would be received (nil, low, moderate, and high)		
Reliability Display:	None		
Cueing:	None		
Display source:	tracking display and signal information		
Independent display variables:	Method for updating target position: manual, in which the subject need to handle the target signals, and automated, in which a tracking system updates target information		
Other Variables / Manipulations:	Task difficulty, i.e., number of targets to be tracked		

Description of Results:
The use of the automated system appeared to be a function of system reliability as well as subjects' confidence in their ability to perform the task manually. However, as task difficulty increased, subjects' reliance on the system increased, regardless of its reliability. More importantly, the results show that when the automated system was performing reliably, subjects' did not notice errors made by the system even with the presentation of feedback and additional time to correct the errors.

Study:	Merlo, Wickens, and Yeh (1999)	**Domain:**	Battlefield
Task:	Target detection, identification, and location	**Bias:**	Attention Trust

Reliability: Degree of precision of the cueing information was highly precise (0°-7.5 from target center), moderately reliable (7.5°-22.5° from target center), or poor precision (22.5°-45° from target center)

Reliability Display: Solid cues signaled highly precise cueing information; dashed cues signaled reduced cue precision

Cueing: Presence or absence

Display source: Far domain imagery presented on the walls of the Cave Automatic Virtual Environment (CAVE) with symbology presented head up using an HMD or head down using a hand-held display, a 3½ inch portable TV.

Independent display variables: World-referenced presentation of head-up imagery vs. non-conformal presentation of head-down imagery

Other Variables / Manipulations: Target expectancy and target priority

Description of Results:

The presentation of highly precise cueing information produced greater benefits for target detection relative to less precise cueing; in fact, target detection performance when cueing information greater than 22.5° from the target was not any different from performance when no cueing information was available at all. However, the presentation of cueing (regardless of its precision) induced an attentional cost exhibited by reduced detections of unexpected but high priority targets when this target was presented with another object which was cued. This behavior was mediated somewhat by reduced trust in the system, when the automation failed unexpectedly, but performance on subsequent trials revealed that trust in the system was quickly regained with the presentation of a few reliable trials.

Study:	Montgomery and Sorkin (1996)	**Domain:**	Generic
Task:	Determine whether information conveyed was signal or noise	**Bias:**	Attention

Reliability: Low, medium, high variance

Reliability Display: Luminance conveyed reliability (white = high reliability; grey = low/equal reliability)

Cueing: Nine information sources (gauges)

Display source: 27cm color monitor

Independent display variables: Reliability (unequal reliability, equal reliability)

Other Variables / Manipulations: Stimulus duration

Distribution of reliability (equal = std dev of each gauge equal to 1; grouped left = std dev of four left elements equal to .75 and std dev of five right elements equal to 1.5; grouped right = std dev of four right elements equal to 1.5 and std dev of five right elements equal to .75; distributed across the even gauges such that the even numbered gauges had a std dev of .75 and the odd numbered gauges had a std dev of 1.5; and distributed across odd gauges so that the odd numbered gauges had a std dev of .85 and the even numbered gauges a std dev equal to 1.3)

Description of Results:

Observers weighted highly reliable sources more than the low reliable sources, but this calibration improved when luminance information was available to convey reliability information.

Study:	Mosier, Palmer, and Degani (1992)	**Domain:**	Air
Task:	Perform preflight inspection.	**Bias:**	Attention Trust

Reliability:
Reliability Display: Electronic checklist which diagnosed the system either as the crew member manually selected each item or automatically when it was displayed.
Cueing: Note from "previous crew" misinforming the current crew that the #1 engine was wearing out and needed to be monitored closely.
Display source: Systems display for a glass cockpit; electronic checklists were displayed in the lower portion
Independent display variables: Checklist information presented on paper, electronically, or not at all (control).
Other Variables / Manipulations: Workload

Description of Results:
The salient cues presented to the user indicated problems with the #1 engine, even though close examination of the system parameters showed that the #2 engine was damaged. Optimally, crews should have left both engines running, but instead, the crews who did this were those who used the less automated procedures, i.e., the paper checklists, which led to more detailed discussions of the problems. Crews who performed the task by memory or those given an electronic checklist cut short their information gathering and shut down the #1 engine given an implicit command by the automation and the compelling misinformation (note) that the #1 engine was failing.

Study:	Mosier, Skitka, Heers, and Burdick (1998)	**Domain:**	Air
Task:	Detect automation failures during simulated flight.	**Bias:**	Attention Trust

Reliability:
Reliability Display: Electronic checklist which diagnosed the system either as the crew member manually selected each item or automatically when it was displayed.
Cueing: Presentation of automated message suggesting an engine fire
Display source: Systems display for a glass cockpit
Independent display variables: Automation failures
Other Variables / Manipulations: Accountability

Description of Results:
Pilots who felt more accountable for their actions monitored the automation more automation more closely than pilots who were not held accountable for the system's performance/failures and detected more automation failures and committed fewer errors. Surprisingly, all pilots detected the engine failure and shut down a correctly functioning engine; subjective reports indicated that pilots used a *combination* of cues to diagnose the fire, when in fact, none were present.

Study:	Ockerman and Pritchett (1998)	**Domain:**	Aircraft maintenance
Task:	Perform preflight inspection.	**Bias:**	Attention Trust

Reliability: Ten faults in the aircraft which were not displayed or presented incorrectly on the computer
Reliability Display: None
Cueing: Checklist items presented in text format only or with text plus picture. (Control condition required that pilots complete the checklist from memory).
Display source: Wearable computer
Independent display variables: display format (memory-based – control, text, picture)
Other Variables / Manipulations:

Description of Results:
Information presented on a wearable computer system was compelling such that pilots did not do as thorough a preflight inspection as pilots without the wearable computer. The effect was more exaggerated with a higher realism picture system, i.e., the greater detail of the pictures makes the system even more compelling on what should be inspected. Consequently, users follow the computer's advice blindly, ignoring their own knowledge or not taking the time to check the accuracy of the computer.

Study:	Riley (1996)	**Domain:**	Generic
Task:	Categorize a character as a letter or number while correcting random disturbances of a marker from target position.	**Bias:**	Trust
Reliability:	90% (working) or 50% (failure)		
Reliability Display:	None		
Cueing:	None		
Display source:	Computer monitor		
Independent display variables:	Automation reliability for the character categorization task		
Other Variables / Manipulations:	Workload Task uncertainty: presentation of characters which were neither letters or numbers Familiarity with automation (students versus pilots)		

Description of Results:
Subjects who were students allocated the categorization task to the automation less than pilots, a result attributed to the fact that pilots were familiar with interacting with automated systems. Students tended to undertrust the performance of the automation and assumed that they could perform the task better.

Study:	Ververs and Wickens (1998a)	**Domain:**	Air
Task:	Fly a simulated path and detect changes in airspeed and perform a landing, watching out for a potential runway incursion.	**Bias:**	Attention
Reliability:	None		
Reliability Display:	None		
Cueing:	None		
Display source:	Evans and Sutherland display		
Independent display variables:	Conventional instrument landing symbology (low guidance) and tunnel-in-the-sky format (high guidance)		
Other Variables / Manipulations:			

Description of Results:
The tunnel-in-the-sky symbology benefited navigation through the specified flight path but the display format was so compelling that pilots seemed to rely solely on the symbology and responded more slowly to traffic in the real world far domain display, reporting that the runway was in sight, as well as the presence of an unexpected runway incursion.

Study:	Wickens and Long (1995)	**Domain:**	Air
Task:	following a pre-specified flight path and executing a landing	**Bias:**	Trust
Reliability:	None		
Reliability Display:	None		
Cueing:	None		
Display source:	Symbology presented on workstation for head-down display with far domain world generated on an Evans & Sutherland computer		
Independent display variables:	head-up versus head-down presentation of imagery conformal vs. non-conformal presentation of guidance information		
Other Variables / Manipulations:	visibility, level of breakout from clouds, presence or absence of a velocity vector		

Description of Results:
An advantage for conformal symbology was present when information was displayed head-up and the far domain was visible. A trade-off was present with the use of non-conformal symbology such that the additional clutter in the forward field of view offset the benefits of reduced scanning. However, the HUD also produced slower response times to the presentation of a far domain unexpected event.

Study: Yeh and Wickens (1998) **Domain:** Battlefield
Task: Target detection, identification, and location **Bias:** Attention
Reliability: 100%
Reliability Display: None
Cueing: Presence or absence
Display source: Far domain imagery presented on the walls of the Cave Automatic Virtual Environment (CAVE) with symbology presented head up using an HMD or head down using a hand-held display, a 3½ inch portable TV.
Independent display variables: World-referenced presentation of head-up imagery vs. non-conformal presentation of head-down imagery
Other Variables / Manipulations: Target expectancy and target priority

Description of Results:
The presence of cueing aided the target detection task for expected targets but drew attention away from the presence of unexpected targets in the environment. This effect was mediated by the display used, such that attentional tunneling was reduced when subjects were using the hand-held display.

Study: Yeh, Wickens, and Seagull (1998) **Domain:** Battlefield
Task: Target detection, identification, and location **Bias:** Attention
Reliability: 100%
Reliability Display: None
Cueing: Presence or absence
Display source: Cave Automatic Virtual Environment (CAVE) with the presentation of imagery simulated with shutter glasses
Independent display variables: World-referenced vs. screen-referenced presentation of cue and heading information
Other Variables / Manipulations: one-eye vs. two-eye viewing
target expectancy and priority

Description of Results:
The presence of cueing aided the target detection task for expected targets but drew attention away from the presence of unexpected targets in the environment. However, analyses support the observation that this effect can be mediated by the use of world-referenced symbology. Displaying symbology to two eyes produced a very slight benefit for target detection when the target was cued.

A1.2 Frame of Reference (FOR)

Study: Olmos, Wickens, and Chudy (1997) **Domain:** Air

Task: Fly to eight different waypoints and monitor for traffic. **Bias:** Attention
Reliability: 100%
Reliability Display: None
Cueing: None
Display source: workstation monitor
Independent display variables: 2D coplanar display, 3D exocentric display, immersed display coupled with a wide view exocentric display
Other Variables / Manipulations:

Description of Results:
Travel benefited from the use of the immersed split screen display, but this benefit was offset by a cost in the time to respond to a traffic threat, in which case the compelling nature of the immersed display captured subjects' attention and they failed to monitor for hazards in the exocentric display.

Study:	Pausch, Burnette, Brockway, and Weiblen (1995)	**Domain:**	Virtual Environment
Task:	navigation through the virtual environment	**Bias:**	Attention

Reliability: 100%
Reliability Display: None
Cueing: None
Display source: hand-held miniature representation of a virtual environment
Independent display variables: tethered hand held display which augments an immersed view of a virtual space
Other Variables / Manipulations:

Description of Results:
An icon presented in the tethered view of the hand-held miniature presentation of an immersive virtual environment captures users' attention such that they cognitively vest themselves into the miniature display rather than using it as only a symbolic representation of their current viewpoint. In fact, users go so far as to orient the viewpoint of the miniature display so that they look over the shoulder of the icon.

Study:	Vincow and Wickens (1998)	**Domain:**	Virtual environments
Task:	Navigation through a 3D visualization of a document database	**Bias:**	Attention

Reliability: 100%
Reliability Display: None
Cueing: None
Display source: workstation monitor
Independent display variables: egocentric, exocentric, and egocentric with an exocentric map
Other Variables / Manipulations:

Description of Results:
Performance on search and judgement tasks benefited from the exocentric presentation of the information database with the egocentric display a close. The presentation of an exocentric view inset in the egocentric display actually hindered performance, capturing subjects' attention and increasing the cognitive demands by requiring integration of two viewpoints.

Study:	Wickens et al. (1999)	**Domain:**	Battlefield management
Task:	Monitor for new or changed enemy activity while answering questions requiring distance judgements, direction judgements, or a count of visible enemy units, and provide a confidence rating in their answer.	**Bias:**	Attention

Reliability: 100%
Reliability Display: None
Cueing: None
Display source: Workstation monitor
Independent display variables: 3D tethered display vs. 3D egocentric display
Other Variables / Manipulations: Subjective confidence in responses

Description of Results:
Use of the egocentric display aided distance judgements were answered best but impaired performance when a count of the number of enemy targets was required due to a difficulty in integrating information across views. The result is similar to attentional tunneling such that soldiers using the immersed display fixated on the information in the forward field of view and did not pan the environment adequately, therefore undercounting the number of enemy targets appearing in positions behind.

A1.3 Scene Detail

Study: Bossi, Ward, Parkes, and Howarth (1997)
Domain: Driving
Task: Track the path of a vehicle while searching for targets in the center and periphery of a driving simulator display
Bias: Attention
Reliability: 100%
Reliability Display: none
Cueing: none
Display source: Driving simulator
Independent display variables: Luminance difference between the information, enhanced in the center of the display, and that in the periphery (high and low)
Other Variables / Manipulations:

Description of Results:
Detection of targets decreased as they were presented farther out into the periphery. More importantly, peripheral target detection and identification performance was degraded when the luminance difference between the information in the center and the information in the periphery was at its highest. The authors hypothesized that this result may be attributable to the richness of the information when it was enhanced therefore benefiting the primary task of tracking; consequently, subjects scanned the periphery less.

Study: Entin (1998)
Domain: Battlefield management
Task: Target detection task performed under time constraints.
Bias: Attention
Trust
Reliability: high (hit rate = .90; false alarm = .10) and low (hit rate = .90; false alarm rate = .40)
Reliability Display: Aiding information highlighted potential targets and presented a numerical rating indicating its confidence in the identification.
Cueing: Targets marked with red squares
Display source: Workstation monitor
Independent display variables: Distortion in image quality
Other Variables / Manipulations: ATR availability
Subjective confidence in responses

Description of Results:
Aided performance better than unaided performance, and aided performance with a high accuracy guidance system was superior to aided performance using a low accuracy ATR system. Subjects' underrelied on the highly reliable system placing too much weight on poor quality visual information.

Study: Kaplan, Kaplan, and Deardorff (1974)
Domain: Architecture
Task: Subjective ratings of functionality
Bias: Trust
Reliability: 100%
Reliability Display: None
Cueing: None
Display source: Workstation monitor
Independent display variables: low or high detailed models of housing developments
Other Variables / Manipulations:

Description of Results:
Ratings showed no preference in the information provided between the high and low detailed models. Additionally, when subjects viewed the actual architectural structure, they felt that the model they had scene, regardless of detail, was an adequate representation.

Study:	MacMillan, Entin, and Serfaty (1994)	**Domain:**	Battlefield management
Task:	(1) Distinguish the presence of a target when it is presented on a blank background, (2) discriminate target shapes from non-target shapes, (3) detect a target in an image containing multiple objects, and (4) detect a target in a degraded image with supplementary numeric information provided to aid detection	**Bias:**	Attention Trust

Reliability:

Reliability Display: Numeric information which indicated the automation's confidence in identifying an object as a target

Cueing: None

Display source: Workstation monitor

Independent display variables: Automation availability
Distortion in image quality

Other Variables / Manipulations:

Description of Results:

Target-recognition performance (hit rate) degraded rapidly over a narrow range of distortion rates, but false alarm rates (likelihood that a target is classified as a non-target) were relatively unaffected. When numerical information was provided to help form a judgment about the object's identity, subjects' hit rate improved but the data still showed biases such that operators overrelied on visual images versus numeric information especially when the image is poor. Consequently, as the quality of the image decreases, operator performance may degrade also because operators fail to rely on numeric data.

APPENDIX 2: PILOT STUDY

To measure the salience of the uncued objects and to determine the equal salience of the unexpected, high priority target (the nuclear devices) relative to the objects (soldiers) they were paired with, pilot subjects were passively moved through the terrain using a scenario in which they were viewing a display from an unmanned ground vehicle. Note that cueing symbology was *not* presented in this study.

A2.1 Subjects

5 subjects (all male) participated in the pilot study; all were members of the US Army or Army National Guard.

A2.2 Apparatus

The experiment was conducted using an immersive virtual reality system, the ImmersaDesk, a drafting table style 4' x 5.5' projection based display. Subjects sat 33" away from the ImmersaDesk at an eye height of 24" so that the display subtended an area of approximately 90º visual angle. Subjects were asked to wear shutter glasses, on which a head tracker was attached.

A2.3 Displays/Tasks

Subjects were passively moved through a terrain environment and asked to scan the display looking for any one of three target objects (soldiers, land mines, and nuclear devices). *Cueing information was not presented in this study*, and in fact, subjects were presented with only those target objects which were uncued in the main study. Consequently, since tanks were always cued, tanks were not presented in the pilot study.

The time during which each target was visible constituted a "trial". Of the target trials, two of the target objects (soldiers and land mines) were presented 86% of the time (6 times each); the nuclear device was presented only 14% of the time, concurrently with an uncued soldier, and was unexpected. Subjects were not told which target to search for but were instructed that reporting the nuclear device took precedence over the detection of the other targets.

Once the target was detected, subjects were asked to report its azimuth using a heading scale, presented in a decluttered format, that appeared when subjects detected the target by pressing the left or right wand button.

A2.4 Procedure

The experiment took approximately 2.5 hours during which subjects were given the instructions for the experiment and then performed the experiment. Subjects were first given a contrast sensitivity test; subjects scoring below 20/20 on this test were dismissed from the study.

Subjects were instructed to pretend that they were battlefield commanders viewing the display from the viewpoint of an unmanned ground vehicle; the display varied in its level of detail. Their task was to find the targets and send information back to their troop regarding the objects' position. Subjects interacted with the display using a wand (with three buttons and a pressure-sensitive joystick) and shutter glasses. Only the buttons were used during the experiment to make responses. The joystick was not used at all. Instead, the speed of travel through the environment was constrained to be 150mph.

Subjects responded to the different tasks using the wand's three buttons. To indicate that a soldier or land mine was detected, subjects pressed the left button on the wand. To indicate that a nuclear device was detected, subjects pressed the right button on the wand. Once the left or right button was selected, a heading scale appeared along the horizon line and subjects were asked to verbally report its location by calling out the target's bearing.

Subjects were presented with 10 paths; each path constituted one block of 14 trials, with a trial defined as the presence of a target object. The experiment consisted of 2 practice blocks and 10 experimental blocks. Each path was presented twice during the experiment: once in high detail and once in low detail. For each level of scene detail, subjects were presented with one practice block, consisting of 7 targets (3 soldiers, 3 land mines, and 1 nuclear device).

A2.5 Performance Measures

The dependent variables collected form the target search task were the distance from the viewer to the target at the time of detection and the accuracy for target detection.

A2.6 Results

The data were examined to measure the salience differences between the three target objects and assess whether the salience of the nuclear device was equal to that of the object it was presented with. Repeated measures of analyses of variance (ANOVAs) were used to analyze the data. The graphs are plotted with averages and error bars representing 95% confidence intervals.

Figure A.2.6.1 reveals the detection distance and accuracy for the three target objects in the low detail terrain.

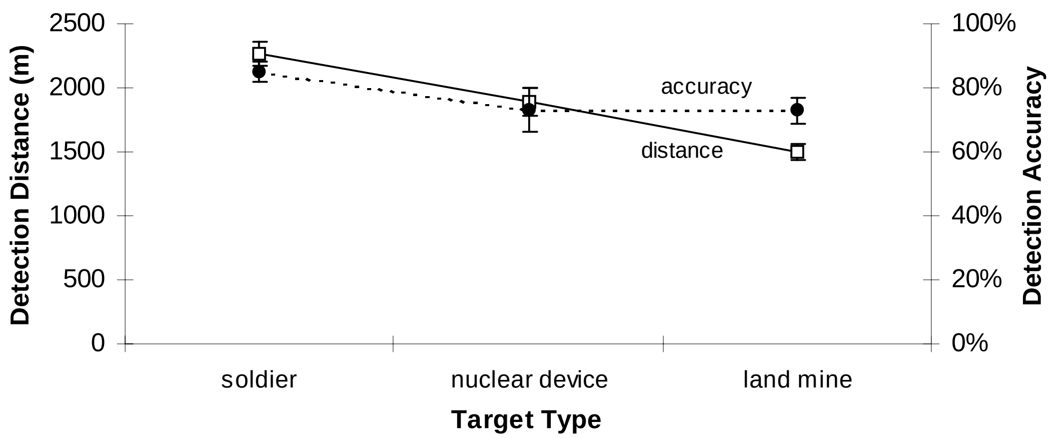


Figure A2.6.1. Detection distance and accuracy in the low detail terrain.

In the low detail terrain, analysis revealed no main effect of target type for target detection distance, $\underline{F}(2,10)=1.35$, $\underline{p}=0.30$, nor accuracy, $\underline{F}(2,20)=0.39$, $\underline{p}=0.69$. More importantly, pairwise comparison revealed no differences due to the detection of the nuclear device on the soldier on either detection distance, $\underline{F}(1,5)=1.37$, $\underline{p}=0.29$, nor accuracy, $\underline{F}(1,5)=0.03$, $\underline{p}=0.88$.

Figure A2.6.2 presents the detection distance and accuracy data in the high detail terrain.

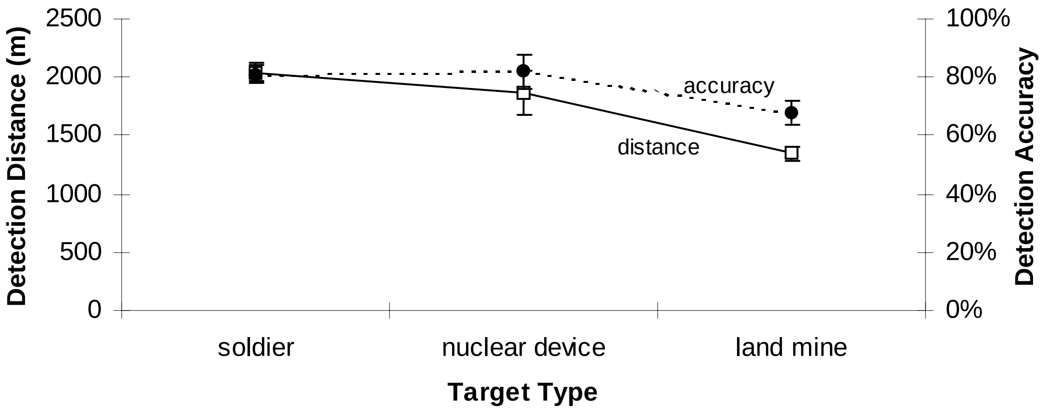


Figure A2.6.2. Detection distance and accuracy in the high detail terrain.

Similar to the detection data in the low detail terrain, there was no main effect of target type on either detection distance, $F(2,10)=0.95$, $p=0.42$, nor accuracy, $F(2,10)=0.03$, $p=0.98$. Pairwise comparison revealed no difference in the detectability of the soldier relative to the nuclear device in terms of detection distance, $F(1,5)=2.05$, $p=0.21$, nor accuracy, $F(1,5)=0.03$, $p=0.87$.

A2.7 Discussion

The result of the pilot study reveal the equal detection of the target stimuli in the low detail and high detail terrain, and more importantly, the equal detection in both distance and accuracy of the (unexpected) nuclear device event relative to the (expected) soldier. In other words, despite the *lower* probability of the appearance of nuclear devices, the data confirm that the nuclear devices *were just as salient* as the soldier with which it was paired.

APPENDIX 3: EXPERIMENT INSTRUCTIONS

Active Navigation

Imagine that you are controlling an unmanned ground vehicle. You will be asked to navigate through a rugged area toward various objectives given a direction of attack. You will move through hilly terrain containing target objects consisting of camouflaged tanks, soldiers, and land mines. Your job is to find these objects and record information regarding their position.

A tunnel representing your desired path will be displayed. <u>You will be responsible for tracking the path as accurately as possible</u>. Try to stay close to the center of the tunnel. At the same time, you will need to scan the environment looking for potential targets. Due to limitations in the sensors which collect the target location information, target objects will be rendered so that they appear somewhat difficult to see in the terrain. High contrast examples of the targets are presented in Figure A3.1.

Passive Viewing

Imagine that you are viewing a terrain display from an unmanned ground vehicle which is heading toward various objectives given a pre-programmed direction of attack. You will travel through hilly terrain containing target objects consisting of camouflaged tanks, soldiers, and land mines. Your job is to find these objects and record information regarding their position. Due to limitations in the sensors which collect the target location information, target objects will be rendered so that they appear somewhat difficult to see in the terrain. High contrast examples of the targets are presented in Figure A3.1.

Figure A3.1. (A) Tank (B) Soldier (C) Land Mine (D) Nuclear Device

In addition to these targets, there will be other objects in the environment, which are not targets and should be ignored. These objects, depicted in Figure A3.2, may be either a truck or a tree.

Figure A3.2. (A) Truck (B) Tree

100% reliable cueing

To help you in the target detection task, you will be presented sometimes with automated target location information; this consists of a lock-on reticle (a box with four crosshairs) superimposed over the target to aid detection. You may assume that the cue is 100% reliable, i.e., whenever the cue appears, it will signal the presence of a target object. However, not all targets will be cued. That is, all target objects (tanks, soldiers, and land mines) have the potential to be cued with the exception of the nuclear devices. The truck and tree will never be cued. Nuclear devices will only be present with other targets. They never appear alone. **If seen, the nuclear device should be reported prior to reporting any other targets.**

Partially Reliable Cueing

To help you in the target detection task, you will be presented sometimes with automated target location information; this consists of a lock-on reticle (a box with four crosshairs) superimposed over the target to aid detection. However, not all targets will be cued. That is, all target objects (tanks, soldiers, and land mines) have the potential to be cued with the exception of the nuclear devices. The truck and tree will never be cued. Nuclear devices will only be present with other targets. They never appear alone. **If seen, the nuclear device should be reported prior to reporting any other targets.**

However, sensor technology is fallible. We are only guaranteed that the sensors will be correct 75% of the time; consequently, at times, the distractor objects (the truck or tree) may be cued. Do NOT report the presence of these objects in the environment, even if they are cued.

You will interact with the world using a wand. A diagram of the wand is presented in Figure A3.3.

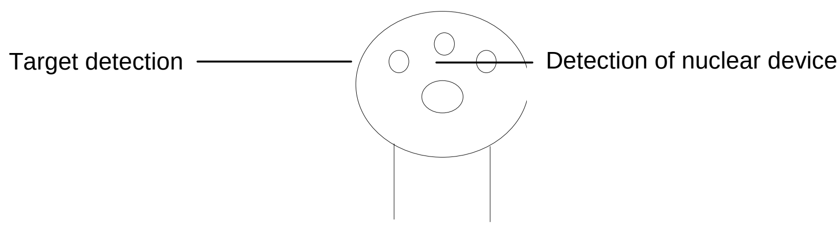


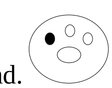
Figure A3.3. The wand.

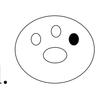
The wand has three push buttons and a pressure-sensitive button. You will use the buttons during the experiment to make responses. The pressure button will not be used at all. You will navigate through the environment by simply rotating the wand in the direction you wish to travel.

THE TASKS

The task you perform consists of two stages: (a) target detection and (b) target location.

Target Detection

In general, to indicate that a target has been detected, press the left button on the wand. 

If the target is a nuclear device, however, press the right button on the wand. 

Target Location

When the wand buttons are selected, compass information will appear on the display, superimposed along the horizon line. You will need to report the target's location by verbally reporting its azimuth.

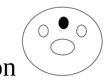
<u>Do not report the presence of any distractor objects.</u>

Local Positioning

Terrain visualization is assumed to be rendered from data recorded and presented by sensors; consequently, the terrain information may not be 100% accurate. Thus, you will also be given a paper map displaying your route. <u>This</u>

map is an accurate depiction of the terrain data. Please report any inconsistencies between the map and what you see in the terrain simulation.

As you move the unmanned ground vehicle through the terrain, its position will be recorded by a global positioning system. As confirmation of its position, four checkpoints, located near distinctive terrain features, are marked on this map and you will need to indicate when the vehicle has passed each checkpoint. To help you in this task, the positioning system will indicate when it believes the vehicle you are controlling is 1000m from the designated checkpoint. You will be able to use this information combined with terrain association in order to report when you have passed the checkpoint.

Press the middle wand button  as soon as you notice that you have passed a checkpoint.

SUBJECTIVE ASSESSMENT

There are two ways to assess the effectiveness of simulator displays and their related tasks. One method is to gauge performance using accuracy scores and response time measures. The second method is to consider the perceived level of workload. Your performance will be measured using both methods throughout the experiment. In order to measure your subjective workload, you will be asked to fill out a NASA Task Load Index (NASA-TLX) at the end of several blocks.

In the experiment, you will view a total of 11 different scenes using two different levels of scene detail. A practice block will be presented to familiarize you with the task.

Thank you for your participation.

APPENDIX 4: QUESTIONNAIRE

A4.1 NASA Task Load Index (NASA-TLX)

The section of the NASA-TLX that you will complete consists of six rating scales. Each scale represents an individual workload descriptor: mental demand, physical demand, temporal demand, performance, effort, and frustration. Circle the place along the index that best describes your workload for the block of trials immediately preceding the administration of the rating scales. Be sure to note the descriptions associated with each of the scales. Performance has “good” on the left and “poor” on the right, while the rest of the scales have “low” and “high” as endpoints. Accompanying the rating scales is a description of the measures. Read the descriptions in order to familiarize yourself with the meanings of the workload descriptors.

Mental Demand: how much mental effort was required to perform the task (e.g., thinking, deciding, remembering)

1 2 3 4 5 6 7 N/A

Low High

Physical Demand: how much physical effort was required to perform the task (e.g., pushing, pulling, reaching, stretching)

1 2 3 4 5 6 7 N/A

Low High

Temporal Demand: how much time pressure you feel to complete the task (e.g., relaxed pace or fast and furious)

1 2 3 4 5 6 7 N/A

Low High

Performance: how successful you feel you were completing the task

1 2 3 4 5 6 7 N/A

Low High

Effort: how hard you work to complete the task

1 2 3 4 5 6 7 N/A

Low High

Frustration: how aggravated or annoyed versus secure or content you feel about accomplishing the task

1 2 3 4 5 6 7 N/A

Low High

A4.2 REALISM[*]

R1. How much did the visual aspects of the environment involve you?

1	2	3	4	5	6	7	N/A
Not at All			Somewhat			Completely	

R2. How compelling was your sense of moving around inside the virtual environment?

1	2	3	4	5	6	7	N/A
Not Compelling			Moderately Compelling			Very Compelling	

R3. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

1	2	3	4	5	6	7	N/A
Not at All			Interfered Somewhat			Prevented Task Performance	

R4. Were there moments during the virtual environment experience when you felt completely focused on the task or environment?

1	2	3	4	5	6	7	N/A
None			Occasionally			Frequently	

R5. What was your overall enjoyment in navigating throughout this environment?

1	2	3	4	5	6	7	N/A
Highly Enjoyable			Mildly Interesting			Not Enjoyable at All	

[*] Adapted from Singer and Witmer (1996).

A4.3 TRUST

T1. How much did you trust the terrain information presented in the computer generated imagery?

1	2	3	4	5	6	7	N/A
Not at All			Somewhat			Very Much	

T2. How much did you trust the information presented in the paper map?

1	2	3	4	5	6	7	N/A
Not at All			Somewhat			Very Much	

T3. [for active navigators] How high was your self-confidence in navigating through the terrain?
[for passive explorers] How high was your confidence in the automatic navigation system which led you through the terrain?

1	2	3	4	5	6	7	N/A
Low			Somewhat confident			Very confident	

T4. How much did you trust the cueing information?

1	2	3	4	5	6	7	N/A
Not at All			Somewhat			Very Much	

T5. How high was your confidence in your ability to locate targets?

1	2	3	4	5	6	7	N/A
Low			Somewhat confident			Very confident	

A4.4 Post-Experiment Questionnaire

How would you rate the fidelity (i.e., the goodness) of the display?

Terrain 1

1	2	3	4	5	6	7	N/A

Not Realistic at All — Somewhat — Very Realistic

Terrain 2

1	2	3	4	5	6	7	N/A

Not Realistic at All — Somewhat — Very Realistic

APPENDIX 5: ANOVA TABLES

A5.1 Detection distance and accuracy: Overall target detection

Detection Distance: Expected, high salience; expected, low salience; unexpected

Source of Variation	*df*	*SS*	*MS*
Interactivity	1	333843.07	333843.07
Reliability	1	2111220.67	2111220.67
Interactivity x Reliability	1	2713115.61	2713115.61
Error	12	5550636.26	462553.02
Realism	1	1440885.15	1440885.15
Realism x Interactivity	1	158348.21	158348.21
Realism x Reliability	1	2219.53	2219.53
Realism x interactivity x Reliability	1	146795.46	146795.46
Error (Realism)	12	2570845.54	214237.13
Target type	2	26074599.87	13037299.94
Target type x Interactivity	2	192729.22	96364.61
Target type x Reliability	2	1823640.40	911820.20
Target type x Interactivity x Reliability	2	342834.24	171417.12
Error (Target type)	24	2758862.98	114952.62
Cueing	1	747425.06	747425.06
Cueing x Interactivity	1	26854.21	26854.21
Cueing x Reliability	1	721667.69	721667.69
Cueing x Interactivity x Reliability	1	25250.71	25250.71
Error (Cueing)	12	2174403.36	181200.29
Realism x Target type	2	45599.35	22799.67
Realism x Target type x Interactivity	2	180497.77	90248.88
Realism x Target type x Reliability	2	16409.21	8204.60
Realism x Target type x Interactivity x Reliability	2	23532.11	11766.05
Error (Realism x Target type)	24	2424091.55	101003.81
Realism x Cueing	1	180328.61	180328.61
Realism x Cueing x Interactivity	1	57689.34	57689.34
Realism x Cueing x Reliability	1	36783.26	36783.26
Realism x Cueing x Interactivity x Reliability	1	265747.05	265747.05
Error (Realism x Cueing)	12	1183992.90	98666.07
Target type x Cueing	2	1410216.77	705108.39
Target type x Cueing x Interactivity	2	243374.15	121687.07
Target type x Cueing x Reliability	2	691966.78	345983.40
Target type x Cueing x Interactivity x Reliability	2	466562.44	233284.22
Error (Target type x Cueing)	24	4136303.63	172345.98
Realism x Target Type x Cueing	2	94667.97	47333.98
Realism x Target Type x Cueing x Interactivity	2	453983.41	226991.71
Realism x Target Type x Cueing x Reliability	2	144436.89	72218.45
Realism x Target Type x Cueing x Interactivity x Reliability	2	15928.12	7964.06
Error (Realism x Target Type x Cueing)	24	2518739.58	104947.48

Detection Accuracy: Expected, high salience; expected, low salience; unexpected

Source of Variation	*df*	*SS*	*MS*
Interactivity	1	0.01	0.01
Reliability	1	0.01	0.01
Interactivity x Reliability	1	0.01	0.01
Error	12	0.83	0.07
Realism	1	1.47	1.47
Realism x Interactivity	1	0.01	0.01
Realism x Reliability	1	0.04	0.04
Realism x interactivity x Reliability	1	0.01	0.01
Error (Realism)	12	0.21	0.02
Target type	2	3.94	1.97
Target type x Interactivity	2	0.08	0.04
Target type x Reliability	2	0.01	0.00
Target type x Interactivity x Reliability	2	0.05	0.02
Error (Target type)	24	0.50	0.02
Cueing	1	0.04	0.04
Cueing x Interactivity	1	0.00	0.00
Cueing x Reliability	1	0.00	0.00
Cueing x Interactivity x Reliability	1	0.04	0.04
Error (Cueing)	12	0.35	0.03
Realism x Target type	2	0.43	0.21
Realism x Target type x Interactivity	2	0.10	0.05
Realism x Target type x Reliability	2	0.08	0.04
Realism x Target type x Interactivity x Reliability	2	0.09	0.04
Error (Realism x Target type)	24	0.35	0.01
Realism x Cueing	1	0.32	0.32
Realism x Cueing x Interactivity	1	0.03	0.03
Realism x Cueing x Reliability	1	0.00	0.00
Realism x Cueing x Interactivity x Reliability	1	0.00	0.00
Error (Realism x Cueing)	12	0.33	0.03
Target type x Cueing	2	0.80	0.40
Target type x Cueing x Interactivity	2	0.05	0.03
Target type x Cueing x Reliability	2	0.07	0.03
Target type x Cueing x Interactivity x Reliability	2	0.00	0.00
Error (Target type x Cueing)	24	0.45	0.02
Realism x Target Type x Cueing	2	0.40	0.20
Realism x Target Type x Cueing x Interactivity	2	0.00	0.00
Realism x Target Type x Cueing x Reliability	2	0.05	0.03
Realism x Target Type x Cueing x Interactivity x Reliability	2	0.01	0.00
Error (Realism x Target Type x Cueing)	24	0.77	0.03

A5.2 Detection distance and accuracy: Detection of Cued Targets

Detection Distance: Cued targets - expected, high salience; expected, low salience

Source of Variation	*df*	*SS*	*MS*
Interactivity	1	32340.29	32340.29
Reliability	1	2704539.04	2704539.04
Interactivity x Reliability	1	683727.73	683727.73
Error	6	2074204.28	345700.71
Realism	1	1798988.22	1798988.22
Realism x Interactivity	1	178220.86	178220.86
Realism x Reliability	1	249110.23	249110.23
Realism x interactivity x Reliability	1	1410577.75	1410577.75
Error (Realism)	6	416733.13	69455.52
Target type	2	93145.44	46572.72
Target type x Interactivity	2	167999.27	8999.64
Target type x Reliability	2	352581.96	176290.98
Target type x Interactivity x Reliability	2	919788.46	459894.23
Error (Target type)	12	533243.42	44436.95
Realism x Target type	2	11862632.04	5931316.02
Realism x Target type x Interactivity	2	22937.76	11468.88
Realism x Target type x Reliability	2	680820.07	340410.04
Realism x Target type x Interactivity x Reliability	2	7564.32	3782.16
Error (Realism x Target type)	12	1019603.90	84966.99

Detection Accuracy: Cued targets - expected, high salience; expected, low salience

Source of Variation	*df*	*SS*	*MS*
Interactivity	1	0.01	0.01
Reliability	1	0.01	0.01
Interactivity x Reliability	1	0.03	0.03
Error	6	0.06	0.01
Realism	1	0.00	0.00
Realism x Interactivity	1	0.00	0.00
Realism x Reliability	1	0.00	0.00
Realism x interactivity x Reliability	1	0.00	0.00
Error (Realism)	6	0.00	0.00
Target type	2	0.02	0.01
Target type x Interactivity	2	0.00	0.00
Target type x Reliability	2	0.02	0.01
Target type x Interactivity x Reliability	2	0.00	0.00
Error (Target type)	12	0.06	0.00
Realism x Target type	2	0.00	0.00
Realism x Target type x Interactivity	2	0.00	0.00
Realism x Target type x Reliability	2	0.03	0.01
Realism x Target type x Interactivity x Reliability	2	0.00	0.00
Error (Realism x Target type)	12	0.09	0.01

A5.3 False Alarm Rate: Distractor Objects

Source of Variation	*df*	*SS*	*MS*
Interactivity	1	0.72	0.72
Reliability	1	2.30	2.30
Interactivity x Reliability	1	0.34	0.34
Error	12	1.68	0.14
Realism	1	0.10	0.10
Realism x Interactivity	1	0.01	0.01
Realism x Reliability	1	0.08	0.08
Realism x interactivity x Reliability	1	0.02	0.02
Error (Realism)	12	0.25	0.02
Target Type	1	0.10	0.10
Target Type x Interactivity	1	0.03	0.03
Target Type x Reliability	1	0.04	0.04
Target Type x Interactivity x Reliability	1	0.11	0.11
Error (Target Type)	12	0.10	0.10
Realism x Target Type	1	0.09	0.09
Realism x Target Type x Interactivity	1	0.04	0.04
Realism x Target Type x Reliability	1	0.02	0.02
Realism x Target Type x Interactivity x Reliability	1	0.00	0.00
Error (Realism x Target Type)	12	0.54	0.04

A5.4 Subjective Ratings

	Interactivity	Active	Passive	
R1	How much did the visual aspects of the environment involve you? (1 = not at all, 7 = completely)	4.44 (0.20)	4.19 (0.53)	F(1,12)=0.10, p=0.76
R2	How compelling was your sense of moving around inside the virtual environment? (1 = not compelling, 7 = very compelling)	4.94 (0.25)	5.44 (0.35)	F(1,12)=0.76, p=0.40
R3	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities? (1 = not at all, 7 = prevented task performance)	3.25 (0.28)	2.88 (0.40)	F(1,12)=0.42, p=0.53
R4	Were there moments during the virtual environment experience when you felt completely focused on the task or environment? (1 = none, 7 = frequently)	5.06 (0.39)	5.56 (0.32)	F(1,12)=0.60, p=0.45
R5	What was your overall enjoyment in navigating through the environment? (1 = highly enjoyable, 7 = not enjoyable at all)	3.06 (0.31)	2.56 (0.46)	F(1,12)=0.68, p=0.43
R6	How would you rate the fidelity (i.e., the goodness) of the display? (1 = not realistic at all, 7 = very realistic)	5.06 (0.28)	4.81 (0.33)	F(1,12)=0.22, p=0.65

	Reliability	100%	75%	
T1	How much did you trust the terrain information presented in the computer generated imagery? (1= not at all, 7 = very much)	4.63 (0.27)	4.25 (0.28)	F(1,12)=0.66, p=0.43
T2	How much did you trust the information presented in the paper map? (1= not at all, 7 = very much)	5.88 (0.38)	5.69 (0.36)	F(1,12)=0.06, p=0.81
T4	How much did you trust the cueing information? (1 = not at all, 7 = very much)	6.00 (0.22)	3.69 (0.28)	**F(1,12)=35.10, p=0.0001**
T5	How high was your confidence in your ability to locate targets? (1 = low, 7 = very confident)	5.25 (0.21)	4.81 (0.33)	F(1,12)=0.88, p=0.37