Display Signaling in Augmented Reality: Effects of Cue Reliability and Image Realism on Attention Allocation and Trust Calibration

Michelle Yeh, MITRE, Bedford, Massachusetts, and **Christopher D. Wickens,** University of Illinois at Urbana-Champaign, Savoy, Illinois

This experiment seeks to examine the relationships among three advanced technology features (presentation of target cuing, reliability of target cuing, and level of image reality and the attention and trust given to that information). The participants were 16 military personnel who piloted an unmanned air vehicle and searched for targets camouflaged in terrain, which was presented at two levels of image realism. Cuing was available for some targets, and the reliability of this information was manipulated at two levels (100% and 75%). The results showed that the presence of cuing aided target detection for expected targets but drew attention away from the presence of unexpected targets. Cuing benefits and attentional tunneling were both reduced when cuing became less reliable. Increasing image realism was compelling but increased reliance on the cuing information when those data were reliable. Potential applications include a cost-benefit analysis of how trust modulates attention in the use of automated target recognition systems and the extent to which increased realism may influence this trust.

INTRODUCTION

Techniques of augmented reality allow the visualization of complex data by superimposing supplementary information, relevant to the task at hand, referenced to the real world beyond. Display enhancements to support operators' interaction in these environments include presenting intelligent cuing information to guide attention throughout the scene and increasing the image reality. The fundamental thrust of this study is to examine the extent to which these two features influence the confidence in, trust in, or reliance on information presented in the augmented reality system.

The focus of the experiment is on *intelligent target cuing,* a form of automation that directs the user's attention to regions of the scene assumed by automation to be important (Entin, 1998; Swennsen, Hessel, & Herman, 1977; Yeh, Wickens, & Seagull, 1999). As an automated attention-directing device, target cuing is a subset of a larger class of automation systems designed to present, enhance, or filter information available to the human user (Parasuraman, Sheridan, & Wickens, 2000) and which also includes highlighting (Fisher & Tan, 1989) and alarms (Parasuraman, Hancock, & Olofinboba, 1997; Sorkin, 1988). Such systems have clear benefits when they work reliably, but at least two general forms of costs have been documented.

First, attentional bias or "tunneling" results when the user becomes focused on the cue, or the object to which attention is directed by the cue, to such an extent that other objects or events known to be of importance are not attended. Such attentional effects have been observed in cuing devices involving aircraft maintenance fault inspection (Ockerman & Pritchett, 1998) and military targets (Merlo, Wickens, & Yeh, 1999; Yeh et al., 1999). Note that these attentional costs can occur even when the automation cuing is 100% reliable

Address correspondence to Christopher D. Wickens, Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign, Willard Airport, 1 Airport Rd., Savoy, IL 61874; cwickens@s.psych.uiuc.edu. **HUMAN FACTORS,** Vol. 43, No. 3, Fall 2001, pp. 355–365.

and, therefore, is always directing attention to a relevant object or event.

Second, the cue may be unreliable, failing to highlight an event, object, or target that it was designed to highlight or falsely highlighting a nontarget. Such imperfect attention-guiding automation has been frequently examined in the context of alarm systems (Parasuraman et al., 1997; Sorkin, Kantowitz, & Kantowitz, 1988; Swets, 1998) and in intelligent diagnosis (Mosier, Skitka, Heers, & Burdick, 1998), but it has received less examination in the context of target cuing (Entin, 1998; Merlo et al., 1999).

Unreliable automation attention direction can induce two opposite forms of cognitive response, characterized by Parasuraman and Riley (1997) as *overuse* and *disuse.* On one hand, overuse, *overtrust,* or complacency characterizes an excessive reliance on the cue and results in a failure to adequately process the raw data (e.g., see Mosier et al., 1998). On the other hand, *undertrust* or disuse characterizes a tendency to ignore automation known to be imperfect. Such a response is particularly prevalent in the case of unreliable alarm systems that generate false alarms (Sorkin, 1988). These two cognitive responses may be coupled if an operator initially overtrusts the automation and is unexpectedly "burned" by its first failure, leading to an inappropriately large swing to mistrust and disuse (Lee & Moray, 1994). Afterward, the recovery of trust may take a long time to occur.

In the present experiment we examine this time course of automation reliance before and after the occurrence of a first failure in a target cuing paradigm. Particular interest is focused on the qualitative nature of cuing and reliability effects as represented by signal detection theory. On one hand, increased reliance characterizes a decrease in the response criterion, beta, such that the operator is more willing to report "target" when the cue indicates that one is present (Swennsen et al., 1977; Wickens, Conejo, & Gempler, 1999). On the other hand, it is possible that the cue heightens the quality of inspection of raw data at the region that is indicated by the cue, thereby increasing the overall sensitivity for a target. Both of these effects (which are not mutually exclusive) will produce cuing benefits, and both might be likely to induce attentional tunneling. However, only the risky beta shift, characteristic of complacency, will manifest in increasing false alarms on those occasions when the cuing is unreliable and a nontarget is cued.

Our interest in this experiment is the effect on sensitivity and response bias of reliable target cuing, and of users' discovery of – and subsequent adjustment to – cuing unreliability. In addition to cuing, our experiment also examines the influence of a second variable that might be hypothesized to influence reliance on automation cuing: the realism of the computer graphics image within which cues are presented.

Few studies have empirically investigated how the distribution of attention within a display is affected by scene detail or how the reality with which the information is displayed biases operators to trust one data source over another. The use of highly realistic visual imagery in environments where the operator must also scan the natural world could implicitly compel attention to be directed away from information that is more critical to the operator's task (e.g., away from an out-the-window view for the pilot) to naturalistic computer-generated imagery, which has the potential to display incorrect data (Theunissen, 1998).

The opposite of enhanced image realism is image degradation, and MacMillan, Entin, and Serfaty (1994) showed that when image quality is poor, operators *undertrust* cuing information imposed to augment the degraded imagery. A third variable, interactivity, was manipulated in this experiment, but its effects, which were generally nonsignificant, will not be discussed in this paper. See Yeh and Wickens (2000) for details.

The goal of the present study was to examine the biases generated by the presentation of cuing (an attentional bias) and cue reliability (a trust bias) and to consider how computer-generated image realism could potentially influence these calibrations of cue reliance. In the current study soldiers flew over computer-simulated terrain as if piloting an unmanned vehicle and were asked to detect camouflaged targets. Sometimes their detection was aided by cuing, but not all targets were cued, and occasionally nontargets were cued. This design allowed us to apply signal detection theory to assess whether the benefits of cuing increased the participants' sensitivity to targets (i.e., improved inspection of the raw data that might underlie a cue) or simply changed their response bias, making them more willing to report any object as a target. It also allowed us to examine how increases in scene realism modulated either of the forgoing effects (i.e., the biases generated by the presentation of cuing and cue reliability, and how scene realism might modulate those effects) in particular leading to an increase in cue reliance, as consistent with the data of MacMillan et al. (1994) and the hypothesis offered by Theunissen (1998).

METHODS

Participants

Of the 16 participants (15 men, 1 woman), who were paid $8.00/hr, 13 were members of the U.S. Army or Army National Guard and 3 were members of the U.S. Marine Corps.

Displays/Tasks

The experiment was conducted using an immersive virtual reality system, the ImmersaDesk (Fakespace, Inc., Kitchener, Ontario, Canada), which is a 122 × 152-cm drafting-table-style projection-based display. Participants sat 84 cm away from the ImmersaDesk at an eye height of 61 cm so that the display subtended an area of 90° visual angle. Participants wore shutter glasses on which a head tracker was attached.

The terrain was developed using geographical data of the National Training Center at Fort Irwin, California. The level of terrain realism was manipulated by varying the number of polygons used to generate the terrain and the degree of texture detail that defined the surfaces. Figure 1a depicts a high-detail scene; Figure 1b depicts a low-detail scene.

The ratio of polygons in the high-detail terrain to those in the low-detail terrain was approximately 5:1. The high-detail scene was textured with a satellite image of the terrain, and a noise image was superimposed that modulated the intensity of the underlying data, hence adding more detail to the scene and enhancing the realism of the terrain. The low-detail scene was simply flat-shaded so that

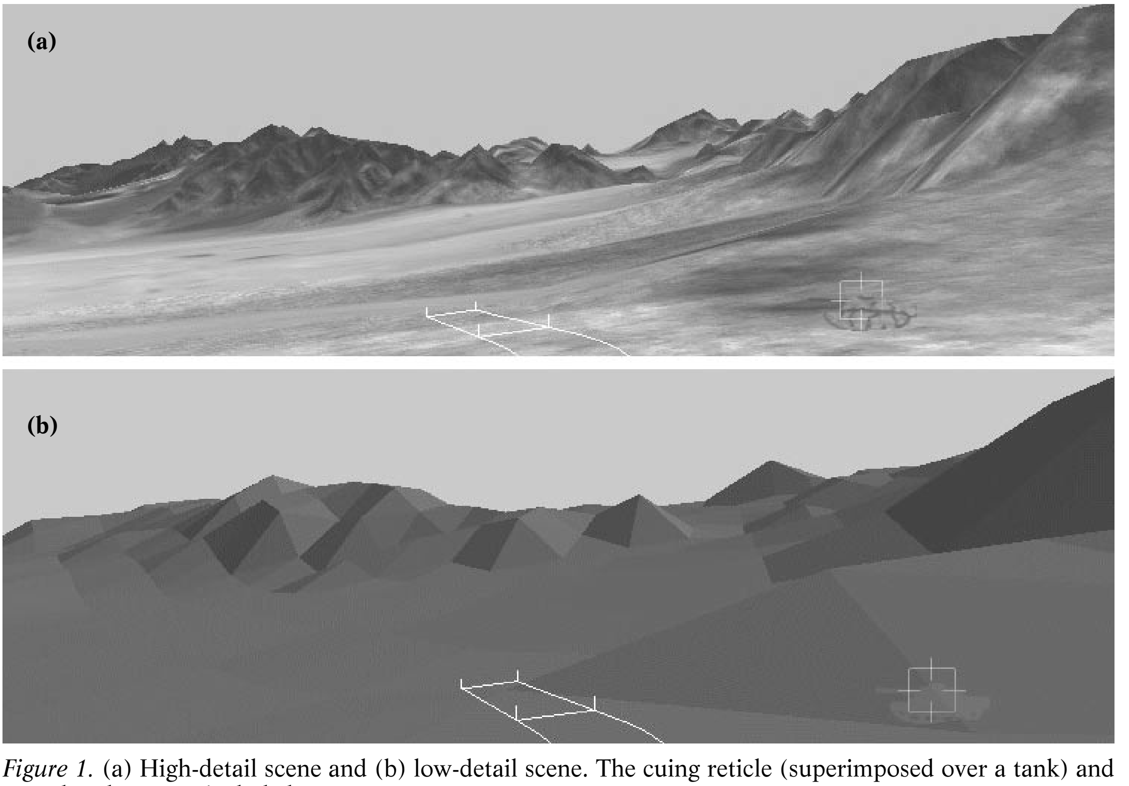

Figure 1. (a) High-detail scene and (b) low-detail scene. The cuing reticle (superimposed over a tank) and tunnel pathway are included.

each polygon was uniformly colored over its surface. The intensity of the color was dependent on the angle between the polygon's normal and a light source; the smaller the angle, the brighter green the polygon would appear.

Six objects, shown in Figure 2, were camouflaged in the terrain. The tank, soldier, land mine, and nuclear device (Figures 2a–2d) served as target objects; the truck and tree (Figures 2e and 2f), which were of shapes that could be confused with the tank and soldier, respectively, were placed in the environment as distractors. The ratio of targets to distractors was 4:1.

In order to degrade the visibility of the objects and thereby force some reliance on the cue, objects in the high-detail terrain were camouflaged (i.e., colored in shades of brown, green, and black) and those in the low-detail terrain were simply silhouetted images shaded green. Note that the examination of image realism is inherently confounded with target detectability because 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. In order to minimize these differences, the intensity of the object was adjusted adaptively at each location so that the contrast ratios between the object and the terrain were similar; this adjustment was based on the performance and subjective assessment of five participants in a pilot study.

Unbeknownst to the participants, except for the nuclear device, only one object could be visible at any given time, and the time during which it was visible was defined as a trial. In the target trials, three of the targets (tank, soldier, and land mine) were presented 90% of the time (30% each) and were therefore considered as "expected"; the fourth (nuclear device) was presented only 10% of the time, within 15° of either a cued or uncued target (always a soldier), and was unexpected but considered to be of high priority. Participants 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. A pilot study with five participants verified that the nuclear device was just as salient as the soldier with which it was paired, a finding consistent with the equivalent size and global shape of the two targets, as shown in Figures 2b and 2d.

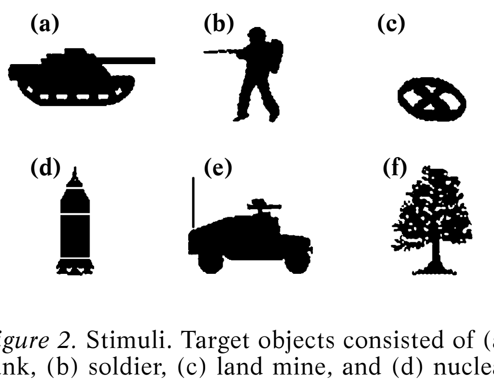

Figure 2. Stimuli. Target objects consisted of (a) tank, (b) soldier, (c) land mine, and (d) nuclear device. Distractor objects were (e) truck and (f) tree.

Symbology consisted of a cuing reticle, heading tape, and tunnel pathway (the cuing reticle and tunnel pathway are presented in Figure 1). When target objects were cued, a reticle consisting of a square with four crosshairs was presented conformally over the target to signal its current lateral and vertical location. The reticle was *not* used to signal the presence of any uncued target. The cuing symbology remained superimposed over the target until it was detected or until it was passed. When cuing was only partially reliable, the cuing symbol sometimes indicated the location of a distractor object, but it usually (75% of the time) signaled a correct target. Participants were instructed not to detect the distractors, even if they were cued.

A heading scale used for azimuth judgment was presented in a decluttered format: When participants indicated they had detected a target by pressing the left or right button on the control stick, the heading tape appeared, conformal to the horizon line. The four cardinal directions were marked on each heading tape. This was used for participants to verbally report the target azimuth. The tunnel symbology marked the path through the terrain (Figure 1).

Experiment Design and Procedure

The manipulations of cue reliability (100% vs. 75%) and interactivity (active vs. passive viewing) were imposed between participants. The manipulations of scene detail (high vs. low) and target type (cued vs. uncued targets, high vs. low expectancy) were imposed within participants. Participants were instructed to assume that they were battlefield commanders seeing the display from the viewpoint of a remotely piloted air vehicle. Their task was target detection, and they were told that automation was available to aid them in this task and were informed as to the level of reliability of the automation. Each path constituted 1 block of trials; a trial can be defined as the presence of an object along the path, whether it was a target or distractor.

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, participants were presented with 1 practice block, consisting of 10 targets, and 5 experimental blocks, each containing a set of 20 targets (6 each of tanks, soldiers, and land mines, and 2 nuclear devices).

Cuing was present on 65% of the trials (i.e., for all tanks, half the soldiers, and half the mines). When cuing was only 75% reliable, all distractors were always cued, and these distractor objects were present (and cued) on 25% of the trials. To foster trust in the cuing information, we first presented participants in the 75% reliability condition with three experimental blocks in which the cuing information was 100% accurate, followed by seven blocks in which the cuing information was only 75% accurate. In general, the data for those first three blocks were not included in the data analyses. However, the first inaccurate cue, which participants encountered on the fourth block, will be critical for the signal detection analysis described later.

After the experiment participants completed a variety of subjective scales, full details of which are reported in Yeh and Wickens (2000).

RESULTS

In each of the following sections the effects of the independent variables (cuing effects, cue reliability, and image realism) and their influence on factors of target detection (i.e., target saliency, expectancy, and cuing) and implicit trust will be discussed. Repeated-measures analyses of variance (ANOVAs) were used to analyze much of the data. The associated graphs are plotted with averages and error bars representing ±1 standard error from the mean. A full discussion of the results, including the muted effects of interactivity, is provided in Yeh and Wickens (2000).

Cuing Effects

The benefits of cuing on target detection performance were examined by measuring the distance at which targets were detected (with greater distance indicating better performance) and calculating the detection accuracy, $1 - P$(target missed). For the purposes of simplifying the discussion of the data, three target classes were formed from the data for the tanks, soldiers, land mines, and nuclear devices.

First, the target objects were classified in terms of expectancy, with the tanks, soldiers, and land mines being expected and the nuclear device being unexpected. Second, prior research had revealed that the land mine, with its smaller visual angle than that of the tank and soldier (Figure 2), was less visible and more poorly detected (Merlo et al., 1999; Yeh et al., 1999). Therefore it was considered the low-salience target. Finally, 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 of uncued targets. Hence the three target classes were (a) expected, high salience (tanks and soldiers); (b) expected, low salience (land mines); and (c) unexpected (nuclear devices).

The detection distance and accuracy data were analyzed using a 2 (reliability: 100% vs. 75%) × 2 (interactivity: active vs. passive) between-subjects × 2 (scene detail: high vs. low) × 2 (cuing: cued vs. uncued) × 3 (target type) within-subject ANOVA. Note that the accuracy data in this ANOVA refers to missed targets, not false classifications of distractors (on the 25% unreliable cuing trials). Figure 3 shows the effects of cuing and target type for the target detection task. Note that in the figure, the unexpected objects were never cued directly; instead, the labels in the figure refer to the status of the object (soldier) with which the nuclear device was paired.

Not surprisingly, the data analysis revealed a significant effect of target type on detection distance, $F(2, 24) = 113.41$, $p < .001$, and accuracy, $F(2, 24) = 94.05$, $p < .001$; the highly

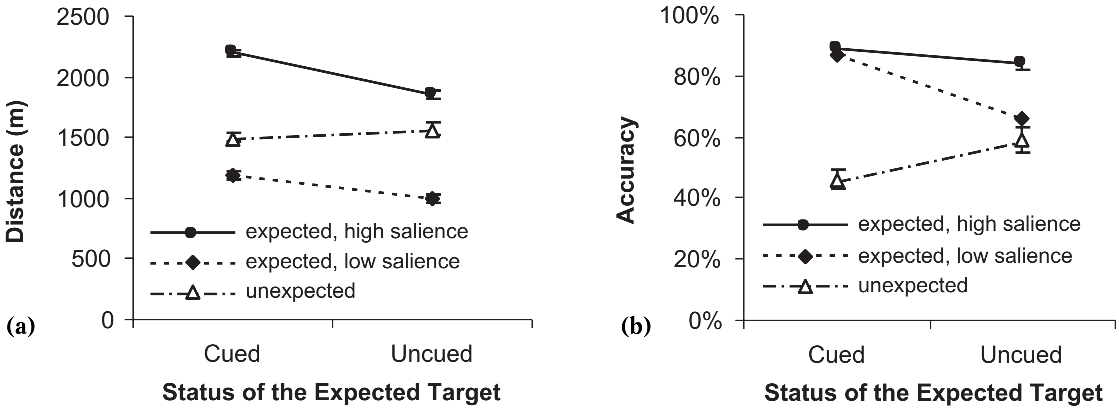

Figure 3. (a) Detection distance and (b) accuracy as a function of cuing.

salient, expected targets were detected at greater distances and with higher accuracy than were both low-salience and unexpected targets. A significant Target Type × Cuing interaction was also observed for distance, $F(2, 24) = 4.09$, $p < .05$, such that the presence of cuing significantly increased the detection distance for the expected objects.

As shown in Figure 3b, the Cuing × Target Type interaction, $F(2, 24) = 23.14$, $p < .001$, was manifest differently in terms of the accuracy data. Although the presentation of cuing did not benefit the detection accuracy of expected, high-salience targets, $F(1, 12) = 0.49$, $p = .50$, it did significantly improve the detection accuracy of low-salience objects by 11%, $F(1, 12) = 42.86$, $p < .001$. More important is the effect of cuing the soldier on the detection of the unexpected, uncued target (the nuclear device) in the same scene. In this case the accuracy data revealed a significant 13% cost to cuing, $F(1, 12) = 6.72$, $p < .05$. That is, replicating effects of attentional tunneling that 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 with an uncued one. This effect was observed despite the explicit instructions that the nuclear weapon was the highest-priority target and the fact that it was always placed within 15° of the cued target.

Partial Reliability

To determine the influence of cue reliability on detection behavior (detection distance and accuracy) for expected objects only, the data were analyzed using the ANOVA model given in the prior section, with two exceptions: (a) only two target classes (high salience and low salience) were included, and (b) for those participants in the partially reliable condition, we excluded the first three blocks of trials, during which the cue was 100% reliable to foster trust in the guidance information.

Unreliability was detrimental to performance (as measured by detection distance), $F(1, 12) = 10.64$, $p < .01$, but was detrimental only on cued trials, $F(1, 12) = 5.69$, $p < .05$; there was no significant effect of reliability on uncued trials, $F(1, 12) = 1.04$, $p = .33$. In fact, when confronted with a partially reliable cue, participants distrusted the cue more and did not use the guidance information when it would have been advantageous to do so. The data reveal no cuing benefits in target detection distance when the cue was only partially reliable, $F(1, 6) = 0.11$, $p = .75$. Note that there were significant cuing benefits to detection distance when the cue was 100% reliable, $F(1, 6) = 14.80$, $p < .001$.

However, this distrust in partially reliable cuing information provided an indirect benefit in that it reduced attentional tunneling, as presented in Figure 4. Examination of the accuracy with which the unexpected but high-priority nuclear target was detected revealed that the significant *cost* for cuing in the 100% reliable condition, $F(1, 6) = 7.67$, $p < .05$, which was seen in Figure 3b, was entirely absent when the cue was unreliable, $F(1, 6) = 1.16$, $p = .32$.

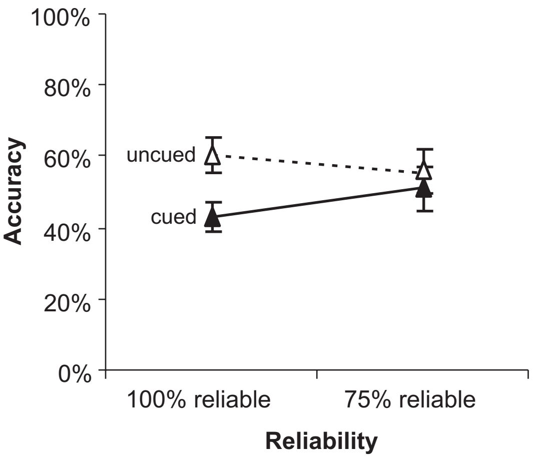

Figure 4. Attention and trust biases: effects of cue reliability on the detection accuracy of the unexpected objects (nuclear devices). Note that the unexpected event was never cued directly; instead, the data presented in the figure are grouped according to the status of the object with which the nuclear device was paired.

To further assess trust in the cuing information when it was unreliable, a 2 (reliability) × 2 (interactivity) × 2 (object type: truck, tree) repeated-measures ANOVA was conducted on the false alarm rate for the distractor objects. 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 cuing of all trucks and trees present in the scene. The analysis revealed a main effect of reliability, $F(1, 12) = 16.46$, $p < .01$; 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 participants' implicit trust in the cuing information, even when they knew it could be unreliable.

Image Realism

An examination of how scene detail influenced target detection performance data suggests improved performance in low-detail scenes: detection distance, $F(1, 12) = 6.73$, $p < .05$; detection accuracy, $F(1, 12) = 82.80$, $p < .001$. Although several steps were taken in pilot work to minimize the differences, an examination of target detection distance and detection accuracy for the uncued objects suggests that a difference was still present for objects of low salience. That is, land mines were detected at a greater distance, $F(1, 12) = 4.48$, $p = .06$, and with significantly greater accuracy, $F(1, 12) = 107.30$, $p < .001$, in the low-detail terrain than in the high-detail terrain.

Signal Detection Analysis

The target detection distance and accuracy measures do not inform as to whether the nature of the errors (i.e., missed targets and false alarms) were the result of a lack in sensitivity or simply an inclination to respond that an object was a target. In order to make such an assessment, signal detection theory was employed to measure how trust in the cuing symbology influenced participants' calibration of how much attention to allocate to an information source, their reliance on this information, and how this reliance was modulated by image realism. Participants' sensitivity and their response criterion (β) were calculated as a function of the accuracy (hit rate) and false alarms. Because of similar salience (i.e., shape and size) between the tank and the truck and between the soldier and the tree, the data from only those four objects were used for the analysis.

Sensitivity was estimated using the nonparametric 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 targets (the tank and the soldier) and FA is the rate of false alarms (Wickens, 1992).

We estimated β using the formula,

$$\beta = e\wedge\{-[z(H)^2 - z(1 - FA)^2]/2\},$$

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 cuing symbology, $P(A)$ and β were calculated for uncued events (which served as our baseline) and compared with corresponding values for cued events when the cuing was 100% reliable and for cued events when the cuing was only partially (75%) reliable. Increased reliance on cuing information would be evidenced by a decrease in β (implying riskier responding based on the presence of the cue). However, this analysis was constrained by the fact that the manipulation of reliability was implemented by the cuing (or noncuing) of distractors.

The implications of the design for the data used in the analysis are shown in Table 1. As Table 1 shows, when cuing information was 100% reliable, all the distractors that were presented were uncued; consequently, a false alarm rate for cued distractors in the 100% reliable condition could not be calculated. Conversely, when cuing information was only partially reliable, all the distractors were cued; consequently, a false alarm rate for uncued distractors could not be calculated with these data. Thus the calculations of $P(A)$ and β for uncued events (a baseline of participants' sensitivity and response criterion when no cuing was available) were estimated solely on the data from participants in the 100% reliability condition.

We wished to estimate how participants' sensitivity and response criterion would change in the detection of cued objects when the guidance information became unreliable – a comparison of reliance when cuing was completely reliable versus partially reliable. Because this comparison could not be made directly, the hit rate supported by 100% reliable cues was estimated using the data from the first three blocks of the partially unreliable condition (i.e., those blocks when participants in the 75% reliable cuing condition were presented with 100% accurate cuing to foster trust in the attentional guidance). The false alarm rate for the 100% reliable cuing was estimated from those participants' responses to the first cued distractor in the fourth block (i.e., the first block in which the cuing "failed" and guided attention to a distractor). This event occurred only once per participant, so a single probability estimate of this value was obtained for all participants. The results for $P(A)$ and β as a function of cuing and reliability are presented in Table 2.

TABLE 1: Availability of Hit (*HR*) and False Alarm (*FA*) Rates for Cued and Uncued Targets as a Function of Reliability

	100% Reliable		75% Reliable	
Cuing	*HR*	*FA*	*HR*	*FA*
Uncued	✓	✓	✓	—
Cued	✓	—	✓	✓

As Table 2 shows, participants were generally less sensitive when cuing symbology was available to aid them in the target detection task than when it was not. More important, as participants' sensitivity decreased, their response criterion shifted so that responses were riskier when they believed that the cuing information was reliable. However, when participants were presented with repeated instances of the automation failure (following Block 4), their sensitivity (and trust in the system) was recalibrated. Sensitivity improved, but not to the level originally seen with no cuing whatsoever. Their response criterion was also adjusted to show a reduced willingness to report a target. However, they were still somewhat guided by the advice of the cue, as witnessed by the lower and therefore riskier β (0.77) setting in the cued than in the uncued condition (2.04).

The use of signal detection analysis provided a means for determining whether differences in the target-background contrast were in fact responsible for the performance differences between targets in the high-detail terrain versus the low-detail terrain. If this were the case, one would expect a corresponding sensitivity difference, with a decreased sensitivity in the high-detail scene relative to the low-detail one. The manipulation of scene realism did have an influence on participants' perceptual experiece, as they found the high-detail scene more involving, $F(1, 12) = 6.55$, $p < .05$, and more compelling, $F(1, 12) = 8.40$, $p < .01$. However, this difference in perception was not reflected by a sensitivity difference (as shown in Table 3, which presents data for the uncued conditions); rather, the data reveal a greater willingness to report a target (a lower β) in the low-detail scene.

We also examined the effect of scene realism on reliance on cuing information, specifically when the cuing information failed the first time. This effect is shown in Figure 5. In examining Figure 5, one can see that the response criterion changed little with realism when targets were cued reliably (the bottom line); however, when the cuing information became less reliable (75%) or unavailable (uncued), the

TABLE 2: Signal Detection Results – Cuing × Reliability

	P (hit)	*P (FA)*
Uncued *P*(*A*) = 0.88, β = 2.04	# targets reported / # targets seen 84%	# distractors reported (false targets) / # nontargets 8%
Cued (100% reliable) *P*(*A*) = 0.64, β = 0.43	# cued targets detected / total cued targets 91%	# cued distractors reported (as targets) / # nontargets (first trial) 63%
Cued (75% reliable) *P*(*A*) = 0.71, β = 0.77	# cued targets detected / total cued targets 89%	# cued distractors reported (as targets) / total cued targets (subsequent trials) 45.5%

data reveal a progressive trend toward a more conservative bias, particularly with a highly realistic scene. That is, as the attentional guidance became less informative, participants were more likely to examine the raw data underlying the cue in the high-detail scene than in the low-detail scene.

This trend toward a more conservative response criterion in the high-detail scene as the cue became less informative was confirmed by subjective ratings (Yeh & Wickens, 2000), which revealed that participants trusted the unreliable cuing information less (and hence were less likely to rely on it) when it occurred in the high-detail scene than in the low-detail one, $F(1, 6) = 9.31$, $p < .05$.

DISCUSSION

The results replicate previous findings of cuing benefits and costs (Merlo et al., 1999; Yeh et al., 1999), but in a different paradigm involving active and continuous navigation. The presentation of cuing information was effective in supporting detection (particularly when it was reliable), and the accuracy benefits of cuing were enhanced when the object that was cued was low in salience.

However, there were two visible costs to the use of this automated attentional guidance: First, the signal detection analysis reveals a shift in the response criterion when cuing was available without a corresponding increase in sensitivity, so that participants were riskier in their responses and likely to overlook errors in the automated cuing system, resulting in a higher false alarm rate. Second, the data reveal an attentional tunneling characterized by decreased detections of the unexpected but high-priority target when it was presented concurrently with

TABLE 3: Signal Detection Results for Uncued Targets – Scene Detail

Realism	*HR*	*FA*	
High detail	78%	8%	P(*A*) = .85, β = 2.11
Low detail	90%	8%	P(*A*) = .91, β = 1.05

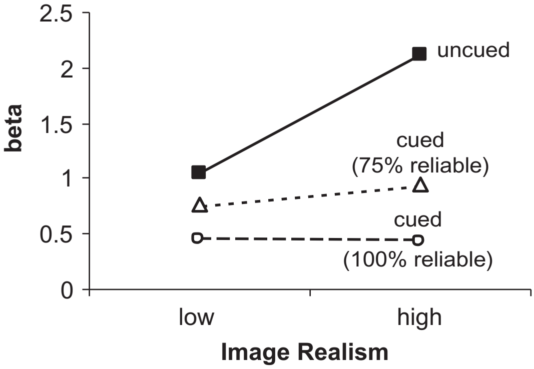

Figure 5. Shifts in β as a function of realism and cuing.

a cued, expected target within 15° of visual angle. This attentional tunneling is one example of the compellingness of automated cuing, which implicitly captures attention and guides it elsewhere at the expense of other objects in the environment. The degree of this attentional tunneling was lowered when the cuing information was unreliable.

In general, partially reliable cuing had three effects: increasing the rate of false alarms, eliminating the benefits of cuing, and decreasing the attentional cost.

Presentation of the cue induced a high level of trust, which, for one group, was unwarranted given the actual precision of the cuing information. Such behavior was predicted and was evidenced by overreliance on the cuing information, as shown by a higher rate of false alarms (i.e., the detection of false targets) when the cue guided attention to the location of a distractor (a truck or tree), a drop in sensitivity, and a risky criterion shift. Subjective ratings clearly reveal that participants trusted the cuing information less when it was only 75% reliable, but their inclination to attend to the cuing information nevertheless rather than the underlying raw data (or target shape) could be attributable to the fact that the target discriminability task was difficult. Consequently, the use of automation increased, even when that automation was not reliable (as predicted by Lee & Moray, 1994, and shown by Wickens et al., 1999).

A consequence of this unreliability was the reduction in target detection benefits for cuing. In fact, when the cue was only partially reliable, the presentation of cuing for valid targets provided no benefits in terms of detection distance relative to uncued objects. Furthermore, the signal detection analysis suggested that participants' sensitivity with unreliable cuing, $P(A) = .71$, was anchored well below that when cuing was unavailable, $P(A) = .88$. Indeed, the loss of reliability appeared to eliminate any of the benefits of cuing insofar as sensitivity and distance was concerned.

The distribution of attention was apparently changed by unreliability in a way that such unreliability decreased the attentional tunneling induced by cuing symbology. When an operator is given multiple sources of information, one information source may be favored at the expense of another (as shown by Wickens, Gempler, & Morphew, 2000), and the reliance that the operator 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 reveal this behavior in detecting the unexpected but high-priority target, as shown in Figure 3; participants 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, participants' loss of trust moderated and broadened their visual search strategies in such a way that they were less attracted by the spatial location of the cue, more likely to scan the surrounding environment, and, hence, more proficient in detecting the unexpected, high-priority events, as shown by the accuracy data in Figure 4. The findings replicate those reported by Merlo et al. (1999) in which participants' breath of attention was widened when their trust in the cue was betrayed.

Our manipulation of realism was effective (as evidenced by subjective data); however, from the standpoint of performance and subjective data, high realism did not engender greater trust as we had hypothesized. Rather, the differences in performance were attributable to the participants' setting of β, in such a way that they were more conservative in their responses (i.e., less willing to report a target) in the high-detail scene (uncued data in Figure 5). More important, though, the signal detection data suggest that realism did influence reliance on the cuing information. As shown in Figure 5, this risky shift was greater in the high-detail scene.

The findings replicate those of MacMillan et al. (1994), whose results suggested that operators would overrely on automation when the image quality was high. However, it should be noted that the magnitude of this shift is totally determined by the baseline, uncued condition. Therefore an alternative implication is that the presence of cuing entirely overrides any influence of scene detail on target detection.

The current research informs theories of attention deployment and the cost-benefit analysis of this information. The results of the study emphasize the need for designers of such cuing systems to carefully evaluate operator reliance on automation. The ability to provide highly precise cuing information – as was implemented in the current study – is subject to constraints such as time delay, incorrect reports from intelligence analysts, limits of sensor resolution, and errors in head-tracking from automation aids, and this imprecision may render the cuing data less than perfectly 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 cuing will increase the likelihood of these errors. The nature of these costs and techniques for preventing such behavior need to be considered in the implementation of any such automated guidance system. At the same time, the results indicated that changes in image realism may modulate cue reliance, implicitly influencing trust and attention deployment throughout the visual scene.

ACKNOWLEDGMENTS

The authors thank Polly Baker, Dave Irwin, Art Kramer, and Nadine Sarter for their suggestions; Ron Carbonari, Albert Khakshour, and Rob Stein for their time, effort, and patience in developing the simulation and data collection software; and David Brandenburg for his assistance in setting up the experimental tasks and in the data collection. Research funding for this project was provided by the U.S. Army Federated Laboratories under Cooperative Agreement DAAL01-96-2-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. government.

REFERENCES

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

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

Lee, J. D., & Moray, N. (1994). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35,* 1243–1270.

MacMillan, J., Entin, E. B., & Serfaty, D. (1994). Operator reliance on automated support for target recognition. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1285–1289). Santa Monica, CA: Human Factors and Ergonomics Society.

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

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.

Ockerman, J. J., & Pritchett, A. R. (1998, May). *Preliminary study of wearable computers for aircraft inspection.* Presented at the International Conference on Human-Computer Interaction in Aeronautics, Montreal, Canada.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics, 40,* 390–399.

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

Parasuraman, R., Sheridan, T., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, 30,* 286–297.

Proctor, R. W., & van Zandt, T. (1994). *Human factors in simple and complex systems*. Boston: Allyn and Bacon.

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

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30,* 445–459.

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

Swets, J. A. (1998). Enhancing diagnostic decisions. In R. R. Hoffman & M. F. Sherrick (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 559–577). Washington, DC: American Psychological Association.

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

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

Wickens, C. D., Conejo, R., & Gempler, K. (1999). Unreliable automated attention cuing for air-ground targeting and traffic maneuvering. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 21–25). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal* 2(2), 99–126.

Yeh, M., & Wickens, C. D. (2000). *Attention and trust biases in the design of augmented reality displays* (University of Illinois Institute of Aviation Tech. Report ARL-00-3/FED-LAB-00-1). Savoy, IL: Aviation Research Laboratory.

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

Michelle Yeh received her Ph.D. in engineering psychology from the University of Illinois at Urbana-Champaign in 2000. She is a member of the D560 HCI and Visualization Group at MITRE in Bedford, Massachusetts.

Christopher D. Wickens is a professor of experimental psychology, head of the Aviation Research Laboratory, and associate director of the Institute of Aviation at the University of Illinois at Urbana-Champaign. He also holds appointments in the Department of Mechanical and Industrial Engineering and the Beckman Institute of Science and Technology. He received his Ph.D. in psychology from the University of Michigan in 1974.

Date received: April 24, 2000
Date accepted: February 15, 2001