DISSERTATION

ENHANCING VISUAL SEARCH PERFORMANCE: INVESTIGATING CUE EFFECTIVENESS, DUAL CUEING, AUTOMATION BIAS, AND ATTENTIONAL TUNNELING IN COMPLEX SEARCH SCENES WITH HEAD-MOUNTED DISPLAYS

Submitted by

Amelia C. Warden

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Anne Cleary
Co-Advisor: Christopher D. Wickens

Dan Graham
Mohammad Arefin

ABSTRACT

ENHANCING VISUAL SEARCH PERFORMANCE: INVESTIGATING CUE EFFECTIVENESS, DUAL CUEING, AUTOMATION BIAS, AND ATTENTIONAL TUNNELING IN COMPLEX SEARCH SCENES WITH HEAD-MOUNTED DISPLAYS

In large complex environments, such as urban driving or flying a plane, human attention may be overloaded, leading to negative consequences when encountering expected and unexpected hazards, like pedestrians crossing the street or a cart on the runway. In such situations, the searcher may benefit from attention cues presented with an HMD. The current experiments address gaps in HMD attention cueing by investigating the effectiveness of different cue properties: cue precision, dual-cueing, cue frame-of-reference, and the impact of imperfectly reliable automation. In all three experiments, participants searched for a routine target (cued or uncued) and an uncued, less expected high priority target. Search efficiency was examined across three different platforms with increasing search field sizes and realism: a static search with a 2D wide-angle desktop display (Experiment 1), a static search presented with an augmented-reality head-mounted display (AR-HMD; Experiment 2), and dynamic search in a 3D virtual reality environment (Experiment 3). Search performance benefited from cueing compared to an unaided search in all experiments. Dual-cueing provided the greatest benefit with the AR-HMD when the searcher's field-of-view (FOV) was constrained by the device's FOV because the searcher benefited from a global cue that indicated which direction they could find the locally cued target. While cueing improved search efficiency, cues showed an overall automation bias, with searchers blindly following incorrect automation. This bias was slightly amplified by the dual cue compared to the single cue. Lastly, there was a trend suggesting automation-based attentional tunneling, where the uncued, less expected high priority target was missed. Overall, attention cueing significantly enhances search performance, particularly with dual cues when targets appear outside of the searcher's FOV. But cueing also introduces an automation bias. These findings have design implications for optimizing automated cueing systems for various platforms to enhance hazard detection in real-world large scenes.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who has supported and guided me throughout my journey of completing this dissertation and my Ph.D. First and foremost, I am profoundly grateful to my advisor, mentor and colleague, Dr. Christopher D. Wickens, for his unwavering support, insightful guidance, and constructive feedback. Your expertise, experience, and encouragement have been invaluable. I would like to thank my other advisors, Dr. Anne Clearly and Dr. Ben Clegg, who have provided insightful guidance and professional development during my work at Colorado State University. I would also like to extend my appreciation to my committee members, Dr. Mohammad Arefin and Dr. Dan Graham, for their engagement, thoughtful critiques and suggestions, and their overall support. A special thank you to the many undergraduate research assistants whom I have had the pleasure and opportunity to advise, mentor, and work with during my Ph.D. Lastly, I am grateful to all the participants, research assistants, and collaborators who contributed to the experiments and studies presented in this dissertation. Your participation has been crucial to the success of this research.

DEDICATION

This dissertation is dedicated to the human factors engineers, domain experts, and designers whose efforts and innovations improve the safety, efficiency, and interactions with technology in our everyday lives. May the findings of this research contribute to the ongoing advancements of attention cueing with emerging technologies and human-automation interaction, ultimately enhancing the design of systems that support and augment human capabilities in complex environments.

As Jane Goodall said, "*What you do makes a difference, and you have to decide what kind of difference you want to make.*"

TABLE OF CONTENTS

INTRODUCTION

Searching for objects or information is ubiquitous in everyday life, from searching for misplaced car keys to scrolling through an email inbox to find a specific message. While there are typically minimal or no consequences from failed searches in routine tasks, more complex searches can results in negative, such as a warehouse worker searching for specific boxes that will be shipped out (Biocca et al., 2007), a pilot searching for airborne hazards (Helleberg & Wickens, 2003), search and rescue personnel looking for a lost hiker, a radiologist searching medical images for broken bones or tumors (Wu & Wolfe, 2019), or a solider searching for an enemy target in the battlefield (Yeh & Wickens, 2001). Searching for critical targets without the assistance of an automation-based aid to guide attention can result in time-consuming and inaccurate searches (Wolfe et al., 2007), which can ultimately have catastrophic consequences. In critical and time sensitive situations, attention cueing, in general, is beneficial (Wickens et al., 2022). Emerging technologies, like those that use augmented reality, have the unique capability of directing attention to critical real-world information by presenting attention cues on either a display overlaying the real-world, such as a head-up display (HUD) or that can embed virtual content that is directly linked to information in the real world, in the case of an augmented-reality head-mounted display (AR-HMD). These newer technologies, particularly AR, are useful for aiding searches in complex environments, like naturalistic scenes or cluttered workspaces, by providing cues that effectively guide visual attention.

Research on cueing using AR does show that automated cueing helps with assembly tasks, in-vehicle navigation (Kim & Wohn, 2011; Morrison et al., 2009), and search tasks (Kumara et al., 2023; Warden et al., 2023; Raikwar et al., 2023). While technologies like AR-HMDs can be effective at aiding searches, they are not without issues. Specifically, automation in general is not perfect (i.e., 100% accurate when cueing a target), and while it may be more accurate than the human alone depending on the complexity of the searches, automation will inevitably make errors (Sebok & Wickens, 2017; Parasuraman & Wickens, 2008), especially the more complex the system is or the more degraded the environmental conditions are in terms of feeding the machine learning algorithm information needed to find the target objects. Research suggests that when automation errs, humans tend to, in general, blindly follow the automation and in the case of cueing aids, tend to follow automation more the more realistic the cue is (Sebok & Wickens, 2017; Manzey et al., 2012; Yeh & Wickens, 2001; Warden et al., 2023).

In addition to automation failures, other factors, such as the placement of information within the display field of view (FOV) and the search environment, pose challenges to attention guidance with augmented reality. Specifically, in real-world, naturalistic 3D environments, like walking along a path on a nature trail, objects can fall outside of the device's FOV. Other important challenges include the design features of the cueing aids themselves and the use of multiple cueing aids simultaneously.

The current research aims to better understand the relationship between multiple factors that impact visual searches in a complex environment when aided by automated attention cues. Specifically, this work seeks to identify (1) the implications of dual-cueing that provides both general and specific spatial information that may benefit searching for targets out of the searcher's field of view, (2) how to best cue attention to information in a complex 3D environment when automated cueing systems error (i.e., identify the wrong target), and (3) the most effective design features that define an HMD cue type, specifically, cue content, precision, frame-of-reference, and the combination of more than one cue.

Visual Search Models and Attention Cueing

Often, searches are completed without the assistance of an aid. The search time during unaided searches, either on a display or in the real-world, can be predicted by the *self-terminating serial search model* (STSS; Neisser, 1964). According to this model, the search time through a set of *N* objects when a target is present is predicted by the formula: $ST = a + \frac{tN}{2}$, where a represents the time associated with making and confirming a response when the target is located, t is the time required to inspect the object and decide it is not the target, and N is the total number of non-target objects in the search field. The division by 2 represents the assumption that, on average, when a target is present it will be found after searching through half of the set size. The self-terminating nature of the model refers to the fact that, when the target is present, the search is terminated once the target is found. In cases where the target is absent, the search is exhaustive because all items must be inspected before determining the target is not present. In such cases, search time can be calculated with the formula: $ST = a + tN$ because all items will have to be examined prior to indicating that the target is absent. Prior research on the STSS model has demonstrated that search time generally increases linearly as the size of the array (*N*) increases, and this slope is twice as steep when the target is absent compared to present due to the exhaustive search required (Sternberg, 1966).

Not all searches are slow or inefficient as predicted by the STSS. For example, pop-out feature searches are preattentive, efficient, and invariant to set size (Treisman & Paterson,1984). Expanding upon this, Wolfe's guided search model (2021) suggests that search efficiencies are on a continuum, shaped by both bottom-up and top-down processes. Specifically, this model integrates how visual attention is directed by both intrinsic features of objects (i.e., bottom-up cues like color and brightness) and the searcher's expectations and prior knowledge of what to search for (i.e., top-down influences). According to Wolfe (2021), these processes dynamically interact to influence search efficiency along a continuum of inefficient to efficient searches. In complex environments, such as the soldier on the battlefield searching for targets, this model explains how the searcher can effectively locate targets during an unaided search despite potential information overload from other objects in the scene and varying levels of object/target saliency.

Collectively, these search models align with our exploration of cue effectiveness in visual search tasks. In the context of an aided search, such as using a cue presented on an HMD, the guided search model (Wolfe, 2021) suggests that cues can enhance visual search performance through a combination of bottom-up and top-down processes. From a bottom-up perspective, cues can be more or less effective depending on how they convey information, such as a cue that highlights a target thereby increasing the salience of that target's spatial location compared to an icon cue that continuously displays the identity of the target but conveys no information about its exact spatial location. From a top-down perspective, cues can provide contextual information that impacts the searcher's expectations of where to look for targets, such as an arrow that orients the searcher to the general direction of a target versus orienting to the precise location. By directing attention more efficiently via both top-down and bottom-up processes, cueing aids can reduce search time and improve search accuracy, especially in a cluttered or complex visual environment. The extent to which we can *combine* these cues to improve search efficiency, such as through the use of two cues (i.e., dual-cueing) has been relatively unexplored and will be a major focus of the proposed research.

Another factor not explicitly accounted for by the previously described models is the concept of *attentional scaling*. This is the idea that attention can be distributed broadly or narrowly over an area of interest (LaBerge & Brown, 1989). Attention scaling impacts performance in everyday activities such as reading, where the attentional focus may shift from an entire page to a single paragraph, or driving, where attention switches from the broader roadway to specific road signs. In the context of visual search, attention can be broadly distributed when searching for a target in a general area, especially in an unaided search where the target is of a certain expectancy and value (i.e., priority relative to other targets). For example, search and rescue personnel looking for a lost hiker broadly distribute their attention when searching large areas of the terrain from a helicopter until they locate signs of the hiker, like a backpack, causing them to narrowly distribute their attention to a specific region on the ground. Expanding upon this logic, attention cues that convey different levels of spatial precision should alter the size of attentional focus, such as a cue that broadly distributes attention to a general area versus one that narrows attention to a specific location (Greenwood & Parasuraman, 2004). If such attentional scaling is correct, this broad or narrow focus of attention should alter the size of the effective search field, where more items will encompass a board than a narrow focus of attention. Furthermore, the combination of general (global) and specific (local) spatial information may better influence search efficiency by directing attention first to the general area while at the same time orienting attention to the specific location. Such redundancy gain should enhance visual search performance and may be contingent upon whether peripheral vision can be exploited during the search.

The STSS model fails to account for factors that impact search, such as the preattentive stage where features are processed in parallel, the similarity between targets and non-targets, and the complexity of the search task (Treisman & Paterson,1984; Wolfe, 2021; Wolfe et al., 2017). Despite these omissions, the model is generally a reasonable approximation and has been confirmed in multiple experiments (See Wickens et al., 2022 for a summary). According to the STSS model, the search time for a perfect cue is predicted to be $ST = a + t$, because, when a cue always locates the correct object, there is only a single object that has to be inspected when determining if it is the target. In cases where the automation is always correct, this hardly requires inspection because attention is guided directly to the correct object. Prior work examining the effects of perfect cueing in an AR-HMD demonstrated that attention cues reduced search time and percent error, on average, by 3.28 seconds (s) and 6.73%, respectively, compared to an unaided search condition (Warden et al., 2022). These findings together with principles from the guided search model and the influence of attentional scaling can better inform insights into effective cue design and how to optimize search efficiency by comparing single and dual cueing systems. The present work seeks to evaluate and compare the effectiveness of different cue types in the context of searching for a hazard within a realistic environment while using a head-mounted display.

Attention Cueing in Augmented Reality

Augmented reality (AR) has the unique capability of overlaying virtual content onto the real-world or embedding virtual content into the real-world. Consider emerging technologies, like the next generation head-up displays (HUDs) or head-mounted displays (HMDs), which can both overlay virtual information onto the real-world and can exploit that AR imagery can be created to establish a one-to-one correspondence between information sources, such as directly overlaying virtual information onto real-world objects or spatial locations. For example, an AR-HMD can link information to the spatial coordinates of real-world information, such as the name of businesses or roadways. Devices like AR-HMDs can be used to guide attention to critical information in the environment, which can assist in many safety-critical and time-sensitive tasks, such as detecting roadway hazards while driving, identifying hostile enemies when moving through a battlefield, or assisting surgeons during medical procedures.

Generic cueing aids that have commonly been used include an arrow cue which points to the location, either general or specific, of a target in a scene, a minimap cue which provides a visual representation of the larger search scene and object locations with the target object highlighted, and an icon cue which conveys identity information about the target in the scene. Figure 1 below shows a generic illustration of each of these cueing aids.

Figure 1. An illustration of an arrow cue (left) that points directly to the target independent of head movement, a minimap cue (middle) that shows a top-down perspective of the search scene and highlights the target, and an icon cue (right) that continuously displays an image of the target on the HMD display, respectively.

While there have been many advancements in the technology of HMDs, such as improved tracking, reduced latency, and weight and comfort of devices, advancements on human factors and performance related issues had slower progress in part due to the cost and availability of the devices, which is now only starting to become more accessible. Prior research has found that AR-based cueing aids are effective for visual search tasks compared to no cueing aid (Kumara et al., 2023; Warden et al., 2023; Warden et al., 2022; Yeh et al., 1998; Yeh et al., 1999). Other studies have found that AR-HMDs resulted in faster task completion and search times (Henderson & Feiner, 2011; Baird and Barfield, 1999; Bauerfeind et al., 2021) and fewer errors (Baird and Barfield, 1999; Tang et al., 2003) compared to other devices (e.g., regular monitor displays, HUD, HDD). However, Wither et al. (2007) found no differences during a visual search task when using an AR-HMD compared to hand-held displays (HHDs) to guide attention, suggesting that effective cueing with AR-HMDs may depend on the context. For example, visual search using an AR-HMD can be somewhat constrained by the limited field of view (FOV) of the device, which typically ranges from 30 to 43 degrees of visual angle. Such a limited field of view can make it difficult to guide attention to a target when searching a large scene, such as a naturalistic area, where information will inevitably fall outside of the immediate FOV of an HMD (Ellis, 2006).

Renner and Pfeiffer (2017) examined how to guide attention to objects outside of the FOV of the device during a manual assembly task where they had to locate objects. Overall, directional cueing (i.e., using an arrow) effectively guided attention to information outside the FOV. While directional cues like arrows effectively orient attention to information inside and outside of the FOV, other traditional cues, such as a highlight cue, lose their effectiveness when a target lies outside of the FOV (Biocca et al., 2007). Furthermore, other traditional cues that provide global directional information, like a map cue, are less effective when finding the precise location of a target compared to other directional cues like an arrow (Biocca et al., 2007; Schinke et al., 2010; Warden et al., 2023). In the context of searching for information outside of the FOV, using a combination of cues, such as an arrow and a highlight cue, may improve cueing effectiveness by quickly orienting the searcher to information outside of the immediate FOV.

While there are benefits to using AR for many real-world tasks (Dey et al., 2018; Jeffri & Rambli, 2021; Tang et al., 2003), presenting content in AR can present several challenges depending on the task. Such challenges include increased display clutter (Warden et al., 2022; Warden et al., 2023), field of view issues associated with device constraints, increased attentional demands, depth perception issues, and issues of accommodation and visual discrepancy (Biocca et al., 2007; Binetti et al., 2021). Four primary issues that arise when cueing attention using AR-HMDs and influence cueing effectiveness that will be addressed in the present work are:

(1) the frame of reference of the cueing aid,

(2) whether the cue conveys spatial or identity information,

(3) the level of precision of the cue, specifically whether the cue conveys global spatial information (i.e., general direction information of a target) or more precise local spatial information (i.e., the exact location of a target), and

(4) the impacts that dual-cueing has on performance compared to a single and no cue condition.

In particular, dual-cueing refers to simultaneously presenting the user with two different concurrent cues to assist in visual search. In the context of the present work, dual-cueing consists of presenting two different cues that convey different levels of spatial precision or presenting an identity (i.e., icon) cue with a spatial location cue. Very few experiments have examined the impacts of such dual-cueing on visual search tasks.

In the following two major sections, the research on these four aspects of the physical properties in cueing that is 100% accurate is described first, followed by the consequences to AR attention cueing that may result when the cue is driven by automation that may be imperfectly reliable, that is, the cue can be wrong. In the work presented here, two forms of cue error are examined: (1) when the cue indicates a non-target as a target, and (2) when the cue fails to indicate a less expected high priority target.

Physical Properties of Attention Cues

HMD cue effectiveness is influenced by physical properties inherent to the cue design. One property is the content of the cue, which distinguishes whether the cue conveys spatial location information or identity information of the target. Cue precision refers to cues that have different granularities of spatial information. They can provide global spatial location information, which conveys the general direction where a target is located in a large search scene, or local spatial information where a cue provides the exact location of the target object whether that object is inside or outside the field of view. This may refer to the specific FOV of an HMD or the less precise FOV defined by wide visual angles (Warden et al., 2022; Poole et al., 2023). The frame-of-reference of a cue, whether it is world- or screen-referenced, also influences how information aligns with the physical environment. Lastly, the use of dual-cueing is another factor that may impact cue effectiveness, where single cues may benefit even more when presented together depending on what and how information is displayed.

Cue Content. Whether the content of a cue is spatial location or identity information is a factor that can impact cue effectiveness. Cues that provide spatial information are generally superior to those providing identity information (Warden et al., 2022; Warden et al., 2023). However, some cues that provide spatial information, such as a highlight cue, while good at directing attention when the object is within the field of view (FOV), can become ineffective when the object is outside of the FOV (Biocca et al., 2007). Prior work also shows that an icon cue resulted in similar levels of accuracy to a minimap (Warden et al., 2023; Warden et al., 2022). Features of the icon cue help guide attention to the correct target (Wolfe, 2021) despite the search taking longer. While spatial cueing has an advantage over identity cueing, there may be circumstances where identity cues are beneficial, such as when the set size of the search array is small, when specific features of the target object help guide attention more efficiently, or when the target and non-targets share a high level of similarity. Such an identity cue can also offload the working memory load to remember what the target object looked like, therefore hedging the tendency to rely on imperfect memory (Ballard et al., 1995). The nature of the identity cue will only represent a minor focus of the current research.

Precision of Cue*.* Another factor that impacts search performance is the extent to which the cue influences attention focus. More specifically, the level of precision the cue provides about the spatial location of the target can influence whether attention is distributed broadly or narrowly (LaBerge & Brown, 1989). Therefore, whether the cue provides global, less previse spatial information compared to local, more precise spatial location information alters attentional focus and therefore the search efficiency. For example, a world-referenced local arrow cue that is displayed, in the real world, at the exact spatial location of an object and points directly to the object, provides very precise, local spatial information compared to a world-referenced global arrow cue that always points to the location of the target object but is presented on the HMD display, and therefore, provides less precise, global spatial information. An even less precise, global spatial cue is the screen-referenced minimap cue that shows the spatial location of the target object from a bird's eye view. Map cues provide spatial location and work well for global scene information but become ambiguous when used for the location of small objects (i.e., the more precise location of a target) in the search scene (Biocca et al., 2007)

Previous work that manipulated attentional focus by looking at the effects of cue size and precision, found that cues directing attention to a global spatial location (i.e., cueing a larger area with less precision) increased response time more than cues directing attention to a local spatial location (i.e., cueing a smaller area with more precision; Greenwood and Parasuraman, 2004). However, this has been shown to depend on cue reliability, where less reliable cues facilitate an expanded attentional focus and therefore are best when they provide more global spatial information than local spatial information (Greenwood & Parasuraman, 1999; Greenwood et al., 1997). Basically, a smaller more precise cue (like an arrow pointing directly at a target) is best for reliable cues, but a larger, less precise cue (like an arrow pointing towards the general direction of where a target is) is best for unreliable cues.

Frame-of-Reference. The frame-of-reference (FoR) for content presented with an HMD can be conveyed in either world-referenced or screen-referenced coordinates. In terms of human information processing, the FoR method used to present a cue on an HMD impacts the cue effectiveness.

Information that is truly considered augmented-reality is world-referenced (a.k.a., world-fixed) information because the content is linked to specific x, y, and z coordinates in the real-world, where imagery changes or updates its position relative to the spatial location that it is linked to in the real-world and as the user changes their viewpoint orientation and position (Billinghurst et al., 1998; Peereboom et al., 2023; De Oliveira Faria et al., 2020). For example, an arrow cue presented on the display in world-referenced coordinates changes its orientation as the head rotates laterally and vertically because the tip of the arrow is linked to the (x, y, z) coordinates of the real-world object. It points directly to the target as though it were tethered to the object in the real-world.

In research on head-up displays (HUDs), **world-referenced** (i.e., AR) imagery is considered *conformal* with the outside world, meaning that the virtual information has a direct link to the real-world counterpart (Wickens et al., 2022; Wickens & Long, 1995; Kim et al., 2016). For example, an artificial horizon line on a HUD moves in tandem with the real world horizon line as the pilot lands the plane. In the context of an AR-HMD, conformal information can be presented on the display, like an arrow cue, or embedded in the real world, like a directional pathway that conforms to the road, where the imagery maintains a direct link with the real-world spatial location. Another example might be a highlight cue that outlines an object in the real-world to signal its importance to the user. AR-HMDs provide the unique ability to link virtual information to explicit spatial locations in the real-world, allowing for the projection of cues at a "contextually relevant location" and offering a "seamless and intuitive mode of information processing" (Bauerfeind et al., 2021; Kim et al., 2018; Robertson et al., 2008; Schankin et al., 2017; Zhao et al, 2023).

Other information presented on an HMD can be fixed relative to the display. **Screen-referenced** (aka screen-fixed or display-fixed) information is considered non-augmented reality information because it is presented at fixed x and y screen coordinates on an HMD display, where information does not change in its location on the HMD as the user changes viewpoint orientation or head position. Such information always moves together with head movements and remains in the FOV (Billinghurst et al., 1998; Lebeck et al., 2017; Peereboom et al., 2023; Fadden et al., 2000). Such cues are screen-referenced and exocentric because they represent information about the cued location independent of the user's position and orientation in space. For example, a minimap cue provides spatial information from an external, top-down perspective (i.e., exocentric viewpoint) where spatial information is inferred based on the location of objects on the map rather than in relation to the user's position or orientation. An icon cue is also screen-referenced because it conveys information about the identity of the target object but not about spatial location.

The literature shows the advantages of world-referenced information over screen-referenced information for various kinds of tasks, including visual search tasks (Billinghurst et al., 1998; Yeh et al., 2003; Warden et al., 2023; Warden et al., 2022). Through perceptual grouping, conformality between the display and the real-world fuses information together to create one object, which helps the user divide attention across the display and the real-world (Wickens et al., 2022; Yeh et al., 1998; Kim et al., 2016; Zhao et al., 2023). For instance, research on attention guidance found that world-referenced imagery improved performance on tasks requiring visual search compared to screen-referenced imagery (Kumara et al., 2023; Renner & Pfeiffer, 2017; Henderson & Feiner, 2011). In some cases, the world-referenced AR imagery also resulted in fewer overall head movements, suggesting AR imagery may be more beneficial and less effortful during search tasks (Henderson & Feiner, 2011). Due to the direct link between virtual content and the real-world, this type of imagery reduces the amount of visual scanning required between information – especially when integration between the two is required (Ellis, 2006). However, this can impose some costs when focused attention is required, such as comparing the identity of an icon cue to a real world potential target, because doing so can be fatiguing depending on whether attention is required for information far away in the real world or information on the display (Ellis, 2006).

Research specific to HUDs shows that conformal (world-referenced) imagery improves performance on non-search tasks (Azuma, 1997; Azuma et al., 2001; Naish, 1964; Lauber et al., 1982, Wickens & Long, 1995) such as flight tracking (Fadden et al., 1998; Fadden et al., 2001) and navigation tasks (Caird et al., 2001; Kim et al., 2018; Boston & Braun, 1996). Conformal HUD imagery also reduces scanning between displayed information and the real-world (Martin-Emerson & Wickens, 1993; Ververs & Wickens, 2000; Wickens & Long, 1994; Boston & Braun, 1996) and reduces negative effects of display clutter (Boston & Braun, 1996; Yeh et al., 2003).

In the context of HMDs, there has been an overall performance benefit (Schmerwitz et al., 2017; Yeh et al., 1998; Yeh & Wickens, 2001; Rusch et al., 2013; Kumara et al., 2023; Renner & Pfeiffer, 2017), as well as reduced attentional tunneling (Schankin, 2017; Schmerwitz et al., 2017; Wickens et al., 2022) when cues are displayed in world-referenced coordinates (i.e., are conformal) compared to screen-referenced coordinates.

The present work aims to expand research examining the impact of cue FoR on realistic visual search tasks to further assess the costs and benefits of conformal, world-referenced cueing compared to screen-referenced cueing. Specifically, this work will look at the influence that world vs screen-reference cues have on performance during both static and dynamic visual search tasks within a realistic environment with a wide field of view. The present work also aims to examine the synergistic potential of combining different cue types, with similar or different FoR, such as a screen-fixed minimap cue presented simultaneously with a world-referenced arrow cue. The juxtaposition of different cue types aims to investigate whether a dual-cueing approach may capitalize on the strengths of each type of cue.

Dual-Cueing*.* The majority of cueing studies have focused on the effectiveness of single cues that direct attention to a single target. However, very few studies have examined the impact that multiple, simultaneous cues have on search performance. In fact, most of the studies that have looked at the concept of multiple cueing have examined the issue in the context of multisensory cues.

Dual-cueing using different sensory modalities has shown mixed findings of its effectiveness. During a visual search task, participants performed worse (error and response time) with a dual-cue consisting of a world-reference arrow that moved in the direction of a target and played a sound when it oriented toward the object compared to a variation of a minimap cue, referred to as the EyeSee360 (Gruenefeld et al., 2018). However, this study is confounded by the fact that the visual cues differed across conditions and the arrow cue had an animation component. In another study, Binetti et al. (2021) examined the effect that visual and auditory cues on a visual search task by using a world-referenced (head-fixed) arrow cue for the visual only condition, and an auditory cue plus the world-referenced arrow cue for the dual-cueing condition. Overall, they found that the combination of visual and auditory cues offset performance costs compared to the visual cue only. While these findings suggest benefits for combined visual and auditory cueing when locating targets, using auditory cues may not always be feasible. For example, a solider in the battlefield navigating through enemy territory cannot draw attention to themselves by using an auditory cue; and the environment itself may be quite noisy, or hearing engaged in vital radio communications. In these cases, using multiple visual cues is one way to mitigate issues with other modalities.

What remains unclear is whether there is a benefit of dual-cueing with visual cues only. Only one study by Hein et al. (2020) has explicitly examined combination cueing with course and fine-grained visual aids during a visual search task. The cueing aids consisted of a world-referenced 2D arrow, the attentional funnel (AF) and spherical-wave (SW). Overall, search performance was faster with the combined cueing conditions (i.e., arrow + AF and arrow + SW) compared to single cueing conditions (i.e., AR, SW, or AF), and the fine-grained cues were better than the course-grain cues. However, the arrow cue was fundamentally different in that it utilized a coarse-grained arrow conveying global location information that switched to an arrow pointing directly at the target, almost making it a combined cue itself rather than explicitly testing the effects of presenting two cues simultaneously. While these results suggest a benefit of dual-cueing, contributions of this work are limited given that they fail to use a baseline, they lack systematic control over the type of content conveyed by the cue and which type of cues are presented together, and the cues were always correct. Little to no research has examined the efficacy of dual-cueing, particularly within the context of a large complex environment and dynamic search scenes, and specifically with cues that present different content (spatial vs identity information) with different frames-of-reference (world vs screen referenced).

Automated Attention Cueing

The previous sections all describe instances where the attention cues were perfect, meaning that they always pointed to the correct target. However, in situations where attention cues are used to orient the user to information in real-time, such as the identification of hazards, the cue depends on automation that is not infallible. Automated guidance systems make inferences that are determined by machine-learning algorithms when examining objects in a search field until the target object is identified. (Warden et al., 2023; Raikewar et al., 2023) These operations are susceptible to failures due to factors such as lighting conditions, pattern complexity of the scene or objects, and the overall reliability of the automated system. That being said, automation attention cues will inevitably lead to an error, such as failing to identify the correct object.

Such inevitable inaccuracies in automation cues bring to light the phenomenon of **imperfect cueing**. This concept was first systematically examined in fundamental visual attention research by Posner (1980) through his cueing paradigm. In Posner's study, participants were cued to the correct target location 80% of the time and the incorrect location 20% of the time (i.e., an imperfect cue). The results found a cueing effect for correct cues, but a cost to performance for incorrect cues. This finding has been replicated throughout the literature, showing a performance benefit for valid compared to invalid cues (e.g., Soret et al., 2019; Yeh et al., 1998; Galster et al., 2001). The repercussions of these errors, such as the additional time required to redirect attention or the potential for missed targets, are critical considerations in the application of automated cueing systems.

Much of the previous literature exploiting Posner's 1978 attention cueing paradigm has been conducted on basic computer screens that lack ecological validity and generalizability to more immersive and complex, 360 degree scenes. Only a few studies have directly examined the effects of imperfect cueing in more complex and realistic environments (Yeh & Wickens, 2001; Warden et al., 2023).
Prior work (Yeh et al., 1999;1998) examined how attention costs and benefits of HMDS during a simulated real-world search task with an HMD for cued expected and uncued and less expected targets that were considered high priority targets. Overall, they found that cueing expected targets was beneficial, but induced cognitive tunneling by directing attention away from uncued and less expected but high priority targets (a nuclear weapon), and more so with more immersive displays (Yeh et al., 2003).

Attentional tunneling is one well known cost of attention cueing and is formally characterized by the allocation of attention to one specific channel of information or task that is longer than optimal and causes a failure to notice information in another channel or failure to perform other tasks (Wickens, 2005). For example, a driver distracted by a cell phone conversation may fail to notice a roadway hazard (Strayer & Johnston, 2001; Wickens, 2005). In the context of attention cueing, attention tunneling is the failure to detect an uncued but high priority target when cued to another secondary target object, a phenomenon observed in prior work (Yeh et al., 2003).

Yeh & Wickens (2001) extended HMD cueing work to examine the impact of perfect and imperfect target cueing with an HMD where participants simulated flying over different terrains and had to find a cued target object and an uncued high priority object. Overall, cueing improved performance but expected targets were found faster and more accurately than a less expected targets. This illustrates the susceptibility to attentional tunneling, particularly when the cue is perfect. While the imperfection of cues hurt performance for cued targets more, they actually reduced the severity of attentional tunneling for uncued targets (Yeh & Wickens, 2001; Yeh et al., 2003), possibility by broadening the attentional scope due to less trust in the unreliable cueing aid. Another key finding from this study was that people identified incorrectly cued targets as the correct target 40% of the time. This suggests that people were sometimes blindly following the automation's advice rather than checking the raw data in the real world to confirm the cued object was the correct target. This phenomenon is referred to as the **automation bias** and is formally defined as a tendency to commit commission errors where people rely on and follow the automation even when it is wrong (Skitka et al., 1999; Manzey et al, 2012), and has been found in other literature as well (Warden et al., 2023).

In summary, these studies show us that cueing is important in enhancing search with an HMD when the cues are perfectly reliable, but can induce attentional tunneling for other uncued targets. Also, concerns arise when an HMD cue is imperfect, such as failing to cue anything (miss) or cueing the wrong objects (false alarm). While imperfection of cueing can help reduce some of the negative effects of attentional tunneling, participants are also susceptible to blindly following the automation as in the automation bias. This literature illustrates the benefits of attention guidance systems, especially when they are perfectly reliable. However, it also highlights the challenges posed by imperfect cueing, an inevitable phenomenon. The three experiments will examine both of these costs of imperfect automation: the automation bias and attentional tunneling.

While the costs and benefits of perfect and imperfect cueing has been established for static searches, little prior research has either extended the effects of imperfect cueing to more realistic search scenes with a wide field of view when cues are presented in augmented-reality, nor has it examined the effects of cueing during a dynamic search, such as moving along a path to find a target.

Current Experiments

Searching complex visual environments where human visual attention may be overloaded or the environment consists of hazards is a difficult task that often results in errors. During these complex searches, which are often safety-critical and time-sensitive tasks, automated attention cueing aids presented on an HMD are beneficial to performance. However, issues can still arise. Specifically, the design features of the cues can impact performance, and automation itself is imperfect. How imperfection interacts with the design features of the cues is not well understood because it is greatly under-represented in the cueing literature. The current set of three experiments aims to understand the influence of automated attention aids on search performance, especially when those aids are imperfect. The work also seeks to understand what design features are best for guiding attention and, in particular, whether dual-cueing can help optimize search in more complex large environments.

In all three experiments, participants completed a visual search task within a naturalistic search scene with a wide field of view. During the task, participants searched for an expected cued target along with a rare, less expected high priority target that was uncued. The attention cues were either perfectly or imperfectly reliable. All three experiments compare the effectiveness of single cues to dual-cues that convey either global spatial location information, precise spatial location information, or identity information.

In Experiment 1, participants completed a static visual search task on a wide-angle 49" inch desktop display. Experiment 2 extends Experiment 1 to an augmented-reality head-mounted display (AR-HMD), where the cue frame-of-reference was manipulated, and the virtually rendered HMD imagery either rotates with the head or is directly linked to the spatial coordinates of the real-world counterpart. The static search task for Experiment 2 consists of a realistic natural scene like Experiment 1 that is projected in the real-world using a virtual display presented by an AR-HMD. Experiment 3 is a fully immersive experience, where participants completed a dynamic visual search task in a realistic three dimensional outdoor environment presented in virtual reality (VR). In this experiment, participants simulated walking forward along a linear path when searching for cued and uncued 3D targets using simulated HMD cues. The rationale for extending the search experiment from an AR-HMD to VR is because simulating a dynamic search task in VR allows for more systematic experimental control than an experiment using an AR-HMD in a real-world outdoor setting. There are many limitations to using the AR-HMD in the real-world, such as technological limitations (i.e., lighting conditions between sunlight and augmented reality overlays, transparency issues, battery life, tracking and calibration issues, the potential for the device to overheat, and connectivity), safety concerns concerning real movement while participants wear an AR-

HMD, weather and temperature conditions, lack of systematic control over environment (i.e., amount of objects in the scene and a limited number of scenarios and contexts that can be tested). Simulating a dynamic search task in VR provides a controllable, flexible, safe, efficient, and ecological valid method to test the implications of cueing during a dynamic search task.

Key hypotheses are listed below, some encompassing two or all three experiments and others examined within a single experiment, in contrasting the results of a specific pair.

- H_1: *Cue Effectiveness***:** Visual search performance will benefit when a target is cued compared to uncued, and this performance benefit will be modulated upward by both cue reliability set sizes.
- H_{1A}: Based on automation literature (Warden et al., 2023), search performance, particularly accuracy, will be better with perfectly reliable cues than imperfectly reliable cues, reflecting the impact of an overall automation bias.
- H_{1B}: Based on prior literature (Yeh et al., 1999; 2001), there will be costs to performance for an uncued, but less expected high priority object due to attentional tunneling. However, this cost to performance will be reduced when the cue is imperfect.
- H_2: *Dual Cueing*: Visual search performance will benefit more from dual-cues than from single-cues. The dual-cueing benefit is predicted to be greatest for cues that provide spatial information, particularly in larger and more complex search scenes (such as those used in E2 and E3).
- H_{2A}: The dual-cueing benefit will be moderated by cue reliability, such that dual cues will hinder performance more than single cues when automation is imperfect, leading to a greater automation bias such that the cue that is best when automation is perfect will result in the worst performance when automation errors. This is referred to as the lumberjack effect, which states that the more compelling or effective automation is when it is perfect, the more it hurts performance when it makes errors (Sebok & Wickens, 2017).
- H_{2B}: Dual-cueing will hurt performance for uncued, high priority targets more than single-cueing due to an increase in attentional tunneling caused by the dual-cue. This is because two simultaneous cues will narrow the attentional focus to a specific location more than a single cue.

- H_3: *Cue Redundancy*: The cue that provides both global and local spatial information is predicted to show a redundancy gain, such that performance is better than either the global or the single cue alone.
- H_{4A}: *Cue Frame-of-Reference*: The frame-of-reference conveyed by the cue will impact performance resulting in greater overall performance for world-referenced cues compared to screen-referenced cues.
- H_{4B}: However, the more compelling and realistic cues (i.e., world-referenced imagery) will also result in greater automation bias and cognitive tunneling. The more realistic the cue the worse its performance will be when the automation errors and the greater degree of cognitive tunneling will be exhibited.

EXPERIMENT 1: STATIC VISUAL SEARCH USING A 2D DISPLAY

Description

The goal of Experiment 1 was to assess the effects of different cue conditions during a wide field of view, realistic search task. Participants completed a visual search task on a wide-angle monitor with a total field of view of 128 degrees. They searched for a cued routine target on each trial, and an uncued and less expected high priority target. To examine overall cueing effectiveness, cueing aids were compared to a no-cue condition, and also between blocks, cues were either 100% reliable or 83% reliable. The point at which human-automation interaction is worse than a human alone occurs when automation is 70% reliable (Wickens & Dixon, 2007). The current work selected 83% reliability for the imperfect automation because it (1) reflects the accuracy of many automated systems and (2) the number of trials for each cue condition impacted the possible reliability levels. Additionally, this experiment assessed the efficacy of single versus dual cues, to examine whether combining cues improves search performance. Results from this experiment will lay the foundation for testing different HMD cue properties during a naturalistic search task consisting of a wide field of view.

Method

Participants. 54 participants were recruited from the psychology research pool at Colorado State University. All participants received course credit in exchange for their participation. The sample size was determined based on the overall cueing effect from previous literature (Warden et al., 2022; Warden et al., 2023), which showed a mean effect size of Cohen's *d* = 0.80. A post hoc power analysis found that 26 participants would be needed to achieve 80% power with an effect size of 0.80 (α = .05, one-sample t-test). However, given that the present research explores relatively unexplored phenomena where the expected effect size is not well established in the literature, a larger sample size was collected to mitigate the risk of underestimating or overestimating the effect size. Furthermore, a larger sample size increases the generalizability of the results because the variability in the sample is a better reflection of the variability of the population (Heidel, 2016). All participants had self-reported normal or corrected-to-normal vision and were screened for colorblindness using an electronic version of Ishihara's test. No demographic information was collected. The experiment was approved by the Institutional Review Board (IRB).

Stimuli and Apparatus. Participants completed the experiment using a 49” wide-angle 2D desktop display with a resolution of 5,280 x 1,440 (Samsung 49” Odyssey Neo G9 Gaming DQHD Quantum Mini-LET Monitor). The cueing aids and the experiment were developed in the game engine Unity (version 2022.1.21f1). The unit of measure in Unity is 1 unity measure, which was 1.375 inches in the present experiment. The search scene images were of naturalistic flat terrain with relatively little foliage. See Figure 2 below for an example of the type of images used for the search scene (top) and how the images were displayed on the wide-angle monitor (bottom).

Figure 2. One example of the naturalistic scene used for the visual search task (top) and the same scene presented on the wide-angle monitor (bottom) with the objects embedded within the scene. Note that the bottom image represents the no cue condition.

Seventeen unique scene images were photographed from a vantage point of approximately 6’ 6” and a downward perspective using a 24mm wide-angle lens. Using a higher vantage point allows for a wider field of view and a larger search scene area, which increases the possible location of objects and mitigates simple search strategies. Furthermore, a larger search field results in smaller visual angles for individual objects, making the search scene sufficiently difficult compared to a smaller scene with limited locations for objects and higher visual acuity of objects. The participant sat 15.5 inches from the monitor to achieve a maximum visual angle of 128 degrees.

The search scene consisted of routine target objects. Realistic images comprising naturalistic items (rocks and logs) and artificial items (metal and plastic) were used for the routine target objects. These different items were used to ensure variety in the search objects and to increase the generalizability of the results. Using Adobe Stock Images, a total of 128 unique images of objects were used, with an average of 32 images for each category (see Figure 3 for example objects).

Figure 3. Examples of each type of routine object category (top) and all five of the less expected, high priority objects (bottom).

Either 12, 24, or 37 of these objects were positioned in the search scene, one of which was randomly designated the target for a given trial. The search scene also consisted of a less expected high priority object. High priority objects could be one of five different improvised explosive devices (IEDs), such as a pipe bomb or a landmine. Images of real IEDs were taken from the Collective Awareness to UXO site (“Explosive Hazards”, 2024), where various types of real-world explosive devices are categorized. The location of the objects in the search scene was based on 64 total possible locations where objects could be positioned. All objects were uniformly distributed across the search scenes based on these possible locations.

All object images were post-processed using Adobe Photoshop's Brightness/Contrast adjustment feature. The contrast and brightness settings for the objects were decreased by 50 units on the adjustment scale, which ranges from -100 to +100. Reducing the contrast and brightness by 50 means that the darker tones in the image were lightened and lighter tones of the image were darkened, creating a more uniform image overall. These adjustments ensured that the visual intensity of the objects matched with the scene image and prevented any individual object from standing out excessively, akin to the pop-out effect.

In order to simulate depth, the apparent size of each image was approximated based on whether the image was located in the foreground, middle ground, or background of the image. The apparent size was computed by dividing the actual size of the object image by the considered distance from the viewer relative to the foreground, middle ground, and background. Since the participant views an image displayed on the monitor, the distance from the viewer to the foreground of the image was considered as one unit, and the distance to the middle ground and background was 2 and 3 units, respectively. The scale for the foreground, middle ground, and background objects was set to 1.4 by 1.4 unity units (1.925 inches x 1.925 inches), 0.98 by 0.98 unity units (1.35 inches x 1.35 inches), and 0.65 by 0.65 (0.89 inches x 0.89 inches) unity units, respectively.

There were 16 different cueing conditions, including a no-cue condition. The attention cues provided either global spatial information (global cue), local spatial information (local cue), or identity information (identity cue) pertaining to the target object. Cues providing global spatial information convey the general direction of where the object is located. The global cues consisted of a global arrow and minimap cue. The global arrow remained fixed in the center of the display and pointed to the general direction of where the target object was in the scene. The minimap cue provided a 2D perspective map showing the locations of each object in the scene, with the target object containing a yellow highlight around the edge of the object. The local cues consisted of a local arrow that hovered directly above the cued target in the search scene, a gaze guidance line cue that projected a line from the center of the display to the cued target, and a yellow highlight cue that outlined the cued target. Note that while the local and global arrow were the same color, the local arrow was always slightly smaller than the global arrow and was always directly above the cued target. The identity cue was an icon cue, which displayed the appearance of the target object at the center of the display.

Participants used either single cues or dual cues. Figure 4A shows an example of each single cue conditions, which provided only one cue at a time. For example, only the global arrow cue would appear when directing the participant to the location of the target object. Single cues contained either global, local or identity information (Figure 4B). Dual-cues consisted of a combination of either a global cue and a local cue, or an identity cue and a local cue (Figure 4C). In both instances, a local cue was always presented. The specific dual-cue combinations were the (1) minimap cue presented with either the local arrow, gaze guidance line, or highlight cue, (2) the global arrow cue presented with either the local arrow, highlight, or gaze guidance line cue, and (3) the icon presented with either the local arrow, highlight, or gaze guidance line cue. The global arrow, minimap, and icon cues were always located at the center of the display. The gaze guidance line cue always originates from the center of the display and is projected to the cued target in the search scene. Each type of cue is illustrated in Figure 4.

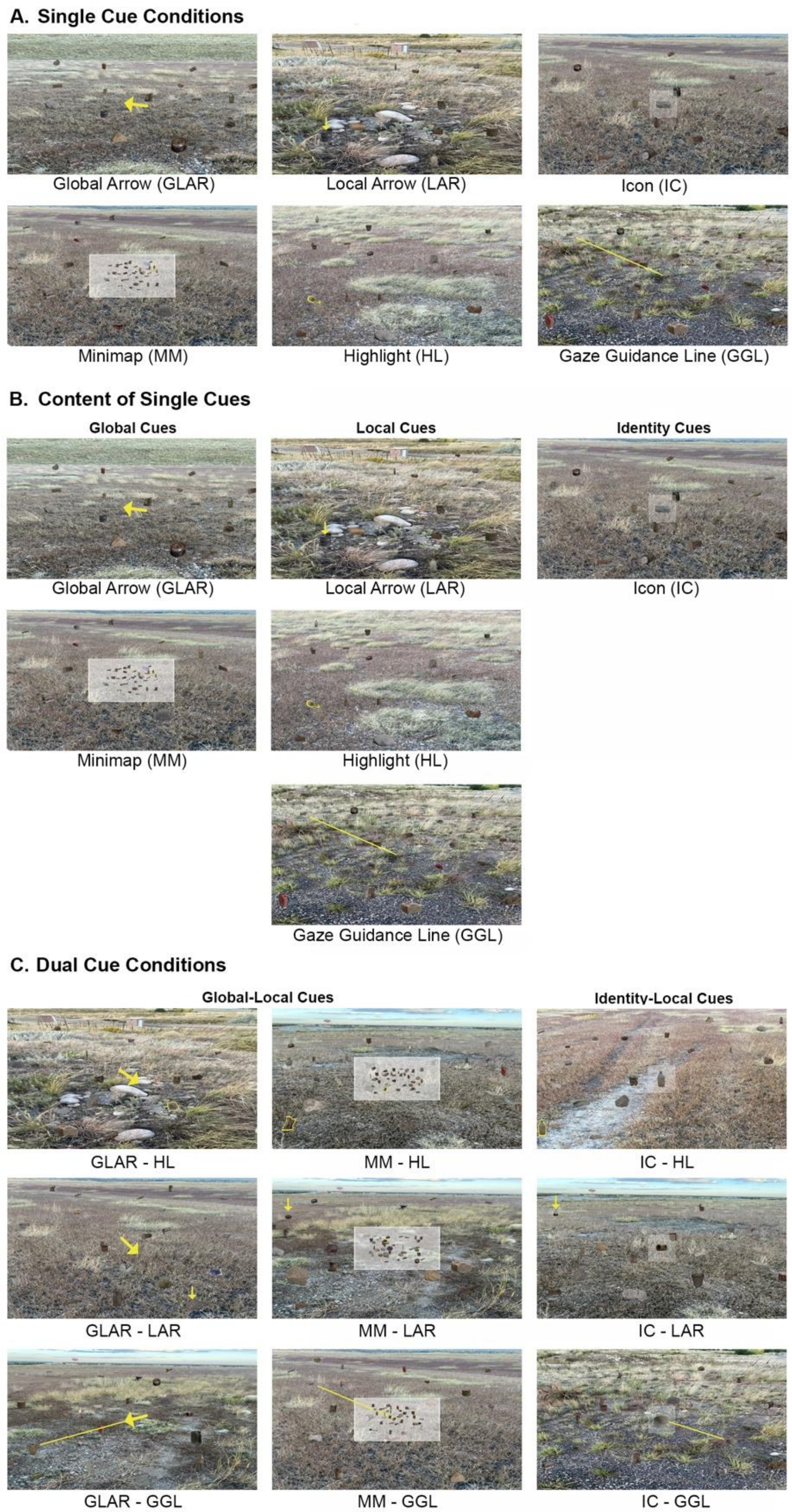

Figure 4. An illustration of each single cue (A), the content conveyed by the single cues (B), and the dual-cues (C) for spatial (global and local) and identity information cues.

In addition to the type of information conveyed by the cue and whether cues were single or dual, the reliability of the cues differed. Cues were either perfectly reliable, meaning they were 100% correct, or imperfectly reliable, meaning that 83% of the time the cue was correct, and 17% of the time the wrong object in the search scene was cued. In the imperfectly reliable condition, the automation identifies a non-target as a target, representing a false alarm. Such an error might happen at an airport security checkpoint, where an automated threat detection system falsely identifies a common object, such as a pack of batteries, as an explosive device. Note that the correct target was always present. Since participants could see the entire screen via peripheral vision, the correct target could be located anywhere in the search scene relative to the incorrectly cued target.

Critically, all cueing conditions exclusively cued only the routine target objects. The high priority object was never cued. This specific experimental constraint was included to assess the effects that cueing has on the detection of other uncued, high priority but less expected targets that may appear in a search scene. For example, imagine a head-up display in a vehicle that has an automated cueing system that uses a machine learning algorithm to identify when pedestrians and other vehicles on the roadway might be intersecting the pathway of the vehicle, but the system is not trained to detect other potential hazards such as a large pothole in the road. While the driver is cued to avoid a collision with a pedestrian or another vehicle, they might miss the pothole, which could result in damaging the vehicle.

Experimental Design. The experiment was a 16 (cue type) x 2 (cue reliability) x 3 (set size) within-subject design. Participants completed two blocks that were counterbalanced by cue reliability such that even numbered participants started with the perfectly reliable cues and odd numbered participants started with the imperfectly reliable cues. Participants completed all cue conditions, including the no-cue, for each reliability block. The cue condition sub-block was counterbalanced within the cue reliability blocks. Within each cue condition sub-block, there were 2 trials for each set size, for a total of 6 trials for each unique cue condition. The 6 trials within each cue condition were randomized. Each reliability block consisted of 96 trials, for a total of 192 trials in the experiment. Participants also completed 2 practice trials for each cue condition for a total of 32 practice trials before they started the experiment. The total duration of the experiment was approximately 60 minutes.

Before the beginning of each trial, participants saw what cue condition they would be using for 1 second, followed by a fixation cross for 300 ms, then an image of the routine target object to study for a total of 3 seconds. The routine target on any trial was one of the naturalistic or artificial objects. For the imperfectly reliable condition, the cue was incorrect 17% of the time, resulting in a cueing error on 1 out of 6 trials for each cue condition. In both the perfectly reliable and imperfectly reliable conditions an uncued, less expected high priority target appeared in the scene on 17% of the trials for each cue condition. The uncued, less expected high priority target was located within 6.875 inches from the routine target object. This equates to the high priority target being within a visual angle of 7.11°, 4.98°, and 3.30° for the objects located in the foreground, middle ground and background, respectively.

Task. The experiment consisted of a 128-degree static visual search task completed on a wide-angle monitor. Participants were asked to search for a routine target object that was cued (except in the no-cue condition). They were also informed that another uncued, target of higher priority might occasionally appear and should always take precedence over finding the routine target. They were still instructed to find the routine target on trials where the less expected, high priority target appeared. Prior to the task, participants read instructions that explained the goal of the task and the different types of attention cues they would use to complete the task (see Appendix A). The instructions also provided information about the imperfectly and perfectly reliable cue conditions. The section for the imperfectly reliable cue condition stated the following:

> "*You will complete trials where the cueing aids may NOT BE PERFECTLY RELIABLE, meaning you may not fully rely on them. Before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.*"

In the perfectly reliable cue condition, participants read the following:

> "*You will also complete trials where the cue is always correct. However, before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.*"

After the search task, participants completed a trust in automation survey, where they answered questions about automation indicating how much they trust the system on a scale ranging from 0 (not at all) to 100% (a great deal). Specific question items for this survey can be found in Appendix D. Lastly, participants answered questions about the cue conditions, such as "*What cue helped you find objects the most?*" and "*Did the single cues or dual cues help your performance most?*" A list of the full questions can be found in Appendix E.

Procedure. Before the experiment, participants gave consent to participate after reading and signing the consent documentation. Participants completed an electronic colorblindness test to screen for any red-green colorblindness deficiencies. Next, they read the experimental instructions and were shown example images of the routine objects and were asked to study the appearance of the five high priority objects. Participants sat in a stationary chair 15.5 inches away from the monitor to ensure an approximate visual angle of 128 degrees. Before starting the main experimental trials, participants completed practice trials to get familiar with each cue condition. Participants used a Logitech M705 Marathon wireless mouse to select the objects during the search task. After they made their object selections, they pressed the 'CTRL' button on the keyboard to continue to the next trial. At the end of the experiment participants completed a trust in automation questionnaire using Qualtrics and were asked a series of questions about the cues they used. The entire experiment lasted approximately 1 hour.

Results

The experimental design consisted of the following main independent variables: cue condition, reliability, and set size. A total of 16 unique cue conditions, including the no-cue, were possible (see Methods section for each unique condition). Cue condition was categorized in different ways depending on the type of analysis, specifically as: (1) cue type: each unique cue condition, (2) dual-cueing: single cues (one cue) versus dual-cues (two cues), or (3) cue precision: global, local, or identity cues.

Before conducting any analyses, the data were examined for outliers using the z-score method and boxplots for both response time and accuracy. The z-score method standardizes the distance of each data point relative to the mean. Data points that are $\pm$ 3 z-scores away from the mean are considered outliers. Based on these methods, two participants were deemed outliers and removed from the analysis. Additionally, four participants did not complete the entire experiment and were removed due to missing data. A total of six participants were excluded from the analysis. All remaining data (N = 47) was analyzed using R Studio.

The assumption of normality was tested using the Shapiro-Wilk normality test, and normality was violated for both response time and error data despite data transformations (*ps* < .05). However, the ANOVA is robust to violations of normality, particularly as sample sizes are within moderate to large ranges where the distribution of sample means begins to approximate the normal distribution as the sample size increases, even if the underlying distribution fails to meet normality (Blanca et al., 2017). Given that the assumption of normality is violated, all reported ANOVA analyses will include the Greenhouse-Geisser (GG) correction to account for such violations. The GG correction is a conservative correction that reduces the degrees of freedom used in the F-test and helps control for Type 1 Errors. Furthermore, simulations have shown that ANOVAs can perform well with moderate violations of normality when sample sizes are moderate to large (Blanca et al., 2017).

All response time data includes both correct and incorrect responses to provide a more complete assessment of the searcher's behavior and performance during the task rather than just the successful outcomes. When analyzing the response time data for routine targets, trials in which participants did not make a response or only selected the high priority target were removed from the analysis to ensure that response time pertained only to the routine targets. Percent error data for the routine targets included all trials.

Descriptive Statistics for Perfect Cue Conditions

Tables 1 and 2 below present cue response times and percent errors collapsed across set size for the perfect cueing condition only. The cue conditions include the icon (IC), minimap (MM), global arrow (GLAR), local arrow (LAR), highlight (HL), gaze guidance line (GGL), and the no-cue condition. As a reminder, global cues (GLAR, MM) are those that provide general spatial location information whereas local cues (LAR, GGL, HL) provide the exact location of a target.

Table 1. Experiment 1: Mean response times, measures in seconds, for each single cue (left column) and the respective dual-cue combinations upper right cells). Standard deviations (SD) are presented in parentheses. These values will be presented graphically in the following text, in the appropriate analyses.

	Single Cues	**IC**	**MM**	**GLAR**	**LAR**	**HL**	**GGL**
IC	**5.18 *s*** (*4.41*)				**2.77 *s*** (*3.25*)	**2.67 *s*** (*1.72*)	**2.46 *s*** (*1.86*)
MM	**4.17 *s*** (*3.33*)				**2.73 *s*** (*2.17*)	**2.52 s** (*1.80*)	**2.86 *s*** (*3.62*)
GLAR	**3.47 *s*** (*2.71*)				**2.72 *s*** (*2.0*)	**2.60 *s*** (*2.0*)	**2.30 *s*** (*1.70*)
LAR	**2.67 *s*** (*1.86*)						
HL	**2.85 *s*** (*2.91*)						
GGL	**2.36 *s*** (*1.72*)						
No-Cue	**5.06 *s*** (*3.98*)						

Table 2. Experiment 1: Mean percent errors for each single cue and the respective dual-cue combinations. Standard deviations are presented in parentheses.

	Single Cues	**IC**	**MM**	**GLAR**	**LAR**	**HL**	**GGL**
IC	**4.96%** (*21.76*)				**1.06%** (*10.28*)	**1.06%** (*10.28*)	**1.06%** (*10.28*)
MM	**4.96%** (*21.76*)				**1.43%** (*11.85*)	**0.71%** (*8.41*)	**1.42%** (*11.85*)
GLAR	**9.57%** (*29.48*)				**1.42%** (*11.85*)	**1.06%** (*10.28*)	**0.00%** (0.00)
LAR	**1.42%** (*11.85*)						
HL	**1.06%** (*10.28*)						
GGL	**0.71%** (*8.40*)						
No-Cue	**13.12%** (*33.82*)						

Figure 5 below shows the mean response time (left) and percent error (right) for each perfectly reliable single cue condition compared to the no-cue condition.

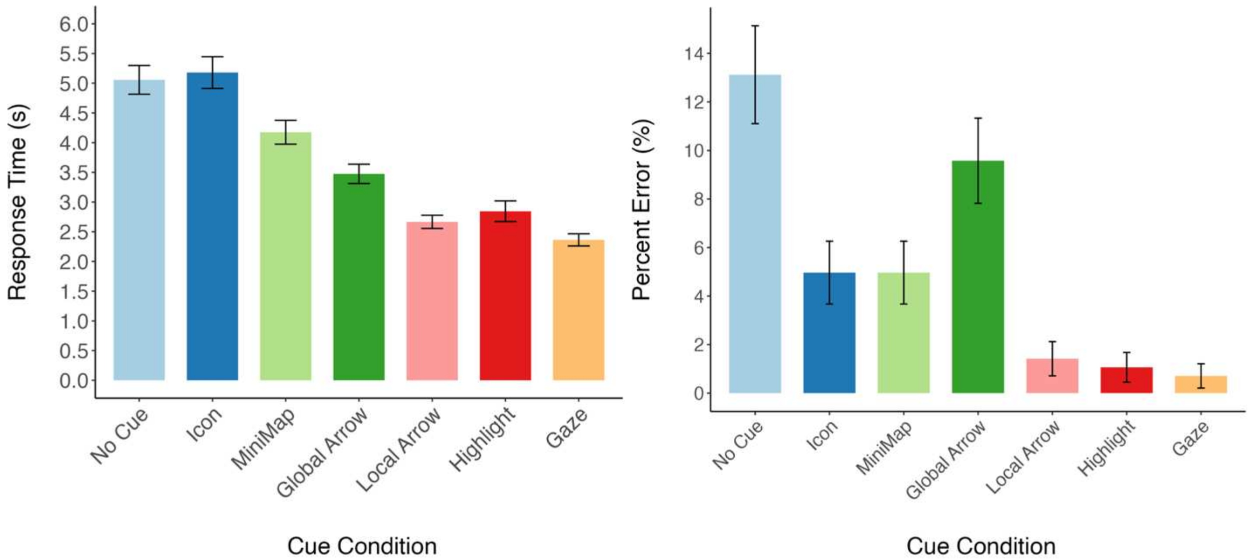

Figure 5. Experiment 1: The mean response time (left) and percent error (right) for all perfectly reliable single cues compared to the no-cue. Error bars represent one standard error of the mean.

A one-way repeated-measures ANOVA was used to examine the overall effect of cue type for both response time and percent error. The results revealed a significant main effect of cue type on response time, $F(3.79, 147.45) = 36.13$, $p < .001$, $\eta_p^2 = 0.44$. The second ANOVA revealed a significant main effect of cue type on percent error, $F(3.07, 141.24) = 12.76$, $p < .001$, $\eta_p^2 = 0.22$. Overall, cue conditions led to faster response times and lower error rates compared to the no cue condition. Specific contrasts will be discussed in more detail in later sections.

Overall Benefits of Cueing by Set Size

A 2 (cue condition) x 3 (set size) repeated-measures ANOVA was used to examine the overall effect of cueing as a function of set size (12, 24, 37). The cue condition was coded as *cue* or *no-cue* (aka *uncued*). The factor *cue* was collapsed across all possible cue types and levels of cue reliability (i.e., perfect, imperfect). Note that Hypothesis 1 predicted an overall benefit of cueing relative to no-cue, and an effect of set size such that search time will increase as set size increases.

Response Time. The mean response time data are shown in Figure 6. The ANOVA revealed that response times were significantly faster when cued ($M = 3.18$ s) than uncued ($M = 5.01$ s), $F(1, 46) =$

102.55, $p < .001$, $\eta_p^2 = 0.69$. The effect of set size was significant, $F(1.72, 82.36) = 27.27$, $p < .001$, $\eta_p^2 = 0.37$, showing that response time increased as set size increased. Critically, the interaction between cue and set size was statistically significant, $F(1.72, 79.31) = 7.31$, $p < .001$, $\eta_p^2 = 0.14$. Overall, response time significantly increased for both cued and uncued searches. However, the slope of this increase is much steeper in the uncued search ($M_{12} = 3.88$ s; $M_{24} = 4.89$ s, $M_{37} = 6.25$ s) than in the cued search ($M_{12} = 2.80$ s; $M_{24} = 3.16$ s, $M_{37} = 3.58$ s). This suggests that the presence of cues significantly improves the efficiency of search time as set size of the search scene increases.

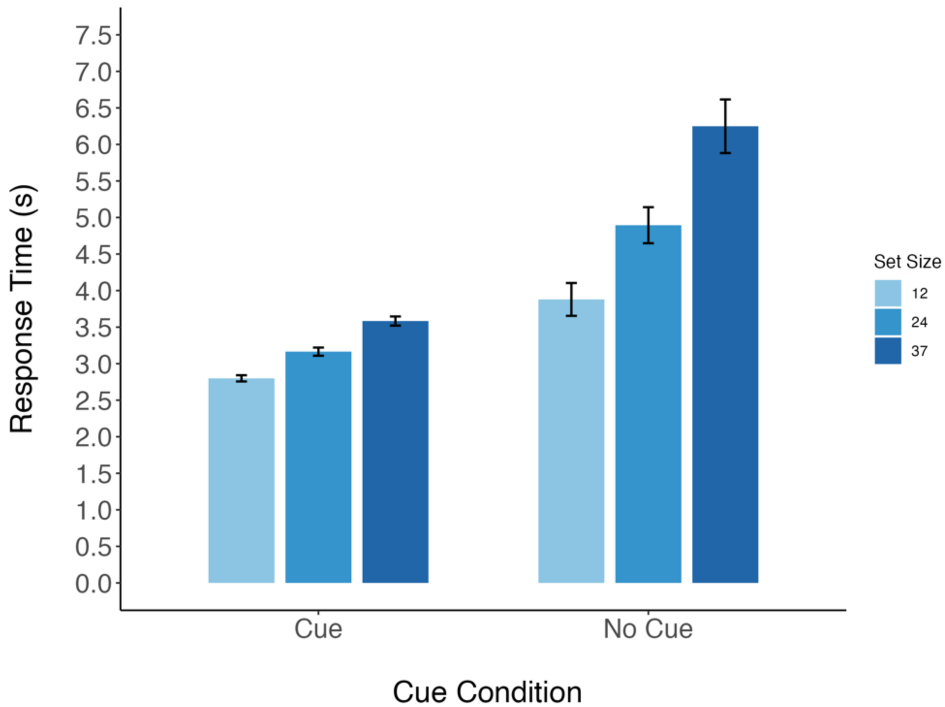

Figure 6. Experiment 1: The mean response time plotted as a function of cue condition (cue, no cue) and set size (12, 24, 37). Error bars represent one standard error of the mean.

Percent Error. Percent error data was analyzed the same way as response time except that percent error was included as the dependent measure. The mean percent error data are shown in Figure 7.

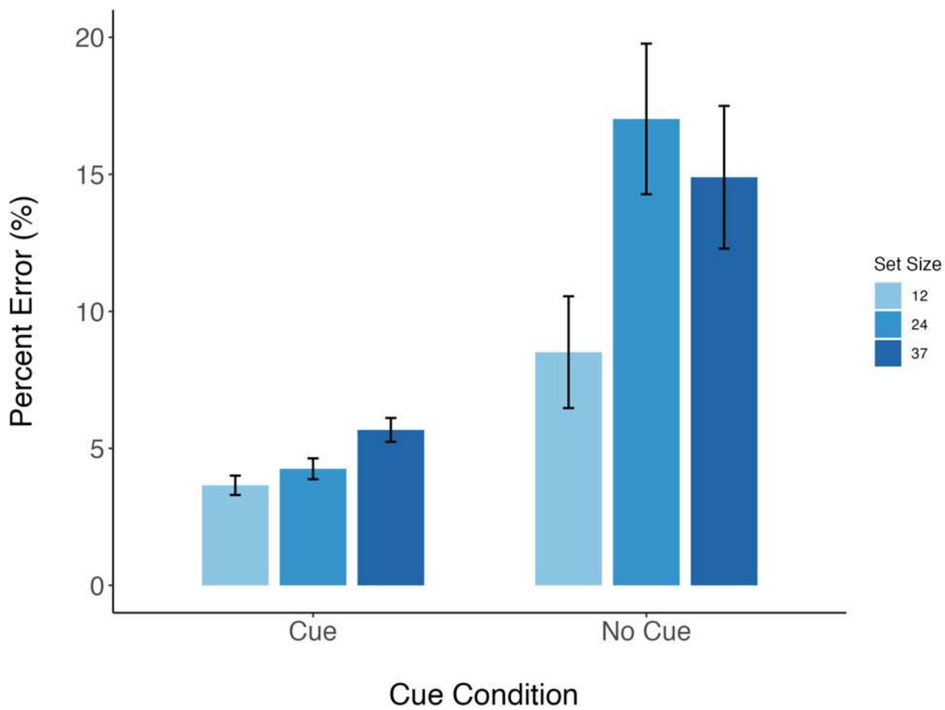

Figure 7. Experiment 1: The mean percent error plotted as a function of cue condition (cue, no cue) and set size (12, 24, 37). Error bars represent one standard error of the mean.

The results show that cues (M = 4.53%) significantly decreased error rates more than the no-cue condition (M = 13.48%), $F(1, 46) = 31.02$, $p < .001$, $\eta_p^2 = 0.40$. The effect of set size was marginally significant, $F(1.79, 82.37) = 3.08$, $p = .051$, $\eta_p^2 = 0.06$, suggesting that error rate increased as set size increased. The interaction between cue condition and set size was not statistically significant, $F(1.81, 83.19) = 2.20$, $p = .116$, $\eta_p^2 = 0.05$.

In summary, search tasks aided by an attention cue improve search performance, allowing for 1.83 seconds (s) faster and 8.95% more accurate searches compared to an unaided search task. While cues effectively guide attention to target objects, enabling quicker and more accurate searches, their effectiveness varies as a function of set size. In general, as set size increases both response time and percent error increases for both cued and uncued searches. However, this effect was more pronounced in the uncued condition, suggesting that the benefits of attention cueing become more critical as the search scene becomes more complex.

Effect of Cue Type and Reliability

The following analyses show the overall effect of reliability collapsed across all cue types and set sizes. Given that the effect of reliability is specific to the automation applied to the cueing aids, the no-cue condition was removed prior to all statistical analyses.

Response Time. A paired *t*-test was conducted to examine the difference between cue reliability conditions. Imperfectly reliable cues (M = 3.51 s) significantly increased response time compared to perfectly reliable cues (M = 3.08 s), $t(46) = 3.07$, $p = .004$, 95% CIs [0.15, 0.73], showing an overall moderate effect of reliability (Cohen's d = 0.43).

Percent Error. Percent error data was analyzed the same way as response time data above. Imperfectly reliable cues (M = 6.93%) significantly increased percent error compared to perfectly reliable cues (M = 2.13%), $t(46) = 7.74$, $p < .001$, 95% CIs [0.03, 0.06], showing an overall large effect of reliability (Cohen's d = 0.96).

Search times were longer and errors were higher when the cue occasionally erred compared to when it was perfect, independent of cue condition and set size. Overall, these findings highlight the importance of understanding the performance decrements that occur with imperfectly reliable automation, especially in environments where quick decisions are required, such as responding to a roadway hazard when driving.

Effect of Dual-Cueing

The following analysis examines the joint effects of cue type and reliability by collapsing cue type into two categories: single versus dual cues. Given that the greatest benefits of dual cueing are expected to be with spatial location cues, the only dual cues considered were those that combined a local cue (highlight, local arrow, gaze guidance line) with a global cue (minimap, global arrow; see Tables 1 and 2). In the imperfect condition for the dual cues, both of the cue types failed 17% of the time. During this failure, both cue types cued the same wrong target object in the scene.

The effect of dual-cueing (single versus dual cues) as a function of reliability was examined using a 2 (cue type) x 2 (reliability) repeated-measures ANOVA for both response time and percent error. The no-cue condition was removed given that reliability is only relevant for the cued conditions. The cue conditions were categorized based on cue type: specifically, as either single cues (i.e., only one cue), dual-cues (i.e., a combination of two cues providing global and local or identity and local information), or no-cue (i.e., the unaided search).

Response Time. The mean response time data are shown in Figure 8. Results show that dual cueing (M = 2.89 s) led to significantly faster response times than single cueing (M = 3.33 s), $F(1, 46)$ = 36.51, $p < .001$, $\eta_p^2 = 0.44$. Also, as reported before, imperfect cues (M = 3.34 s) significantly degraded response time more than perfectly reliable cues (M = 2.84 s), $F(1, 46) = 11.71$, $p = .001$, $\eta_p^2 = 0.20$. The interaction between cue type and reliability was not significant, $F(1, 46) = 0.35$, $p = .56$, $\eta_p^2 < .01$. Overall, imperfect cues impacted dual and single cueing equivalently.

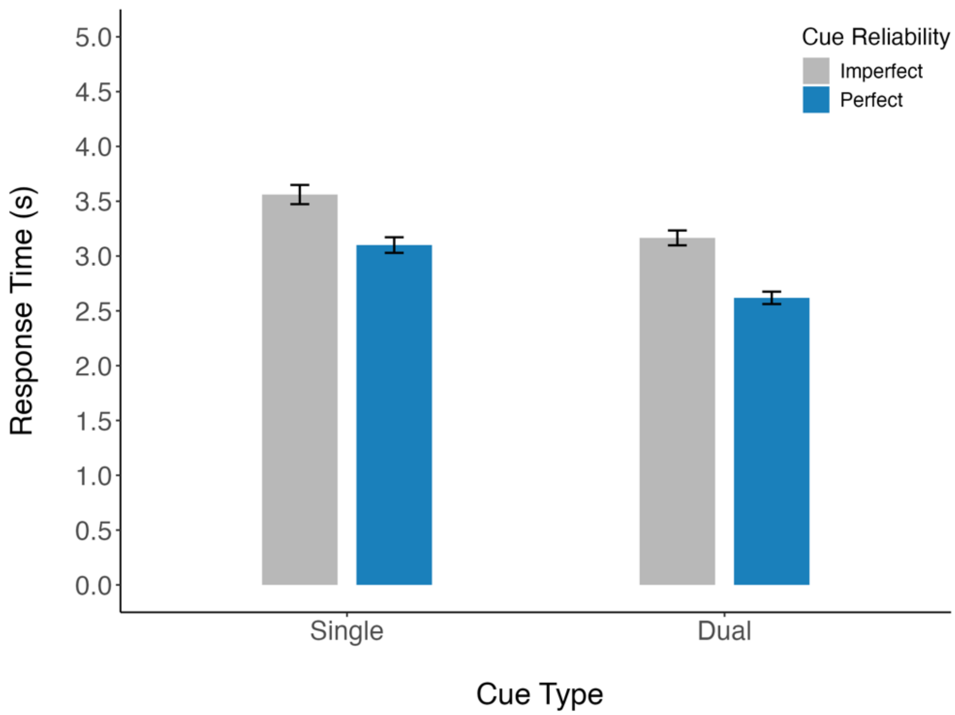

Figure 8. Experiment 1: The mean response time for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

Percent Error. Percent error was analyzed the same way as response time except that percent error was the dependent measure. The mean percent error data are shown in Figure 9.

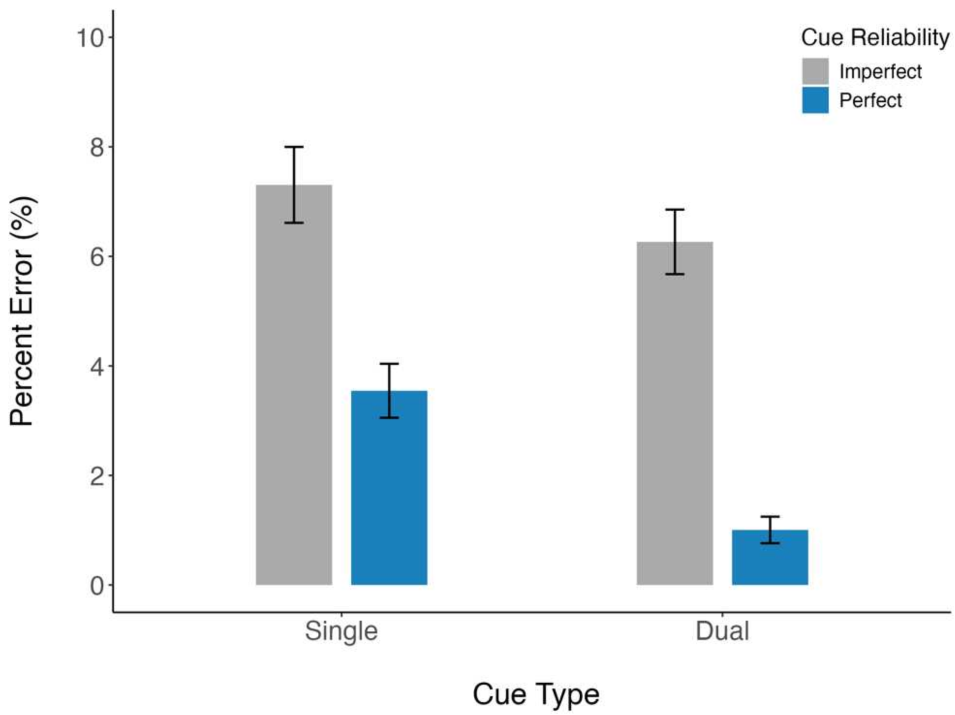

Figure 9. Experiment 1: The mean percent error for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

The results show that dual cueing (M = 3.63%) significantly reduced errors compared to single cueing (M = 5.43%), $F(1, 46) = 8.65$, $p = .005$, $\eta_p^2 = 0.16$. Additionally, as shown before, imperfect cues (M = 4.90%) led to significantly more errors than perfect cues (M = 1.36%), $F(1, 46) = 36.76$, $p < .001$, $\eta_p^2 = 0.44$. Like response time, the interaction was not significant, $F(1, 46) = 2.34$, $p = .133$, $\eta_p^2 = .05$.

Overall, the degrading effect of imperfect cueing was similar for both response time and accuracy with both single and dual cues, indicating that dual cueing did not hinder search time to any greater extent when it is imperfect.

The Automation Bias: Correct vs Incorrect Automation Cueing

The effect of reliability (Figure 9) showed that the largest increase in errors occurred with the imperfect cue condition, suggesting an automation bias. To look at the automation bias, only performance in the imperfect condition will be examined, and this data will be separated into two categories of *automation performance*: when the automation was correct versus incorrect. Trials where only the less expected high priority target was selected were removed before the analysis because the analysis only examines the automation bias for the cued routine objects. In addition, the only dual cues included were those combining global and spatial location information.

While both response time and percent error results are presented, focusing on the error rate in the imperfect cue condition to assess automation bias more directly reflects the user's reliance on the cues, especially when they are misleading. Automation bias involves users trusting automated systems even when they are wrong. If users consistently follow erroneous cues, this indicates a strong automation bias: people blindly follow the advice of the automation and fail to check the raw data in the real world. Response time might not capture this as effectively because it can be influenced by various factors, including decision-making speed and hesitation (i.e., second guessing themselves), not just trust and reliance on automation. Thus, error rates provide a clearer measure of the extent to which users are misled by unreliable cues.

Response Time. The mean response time for the automation bias is shown in Figure 10.

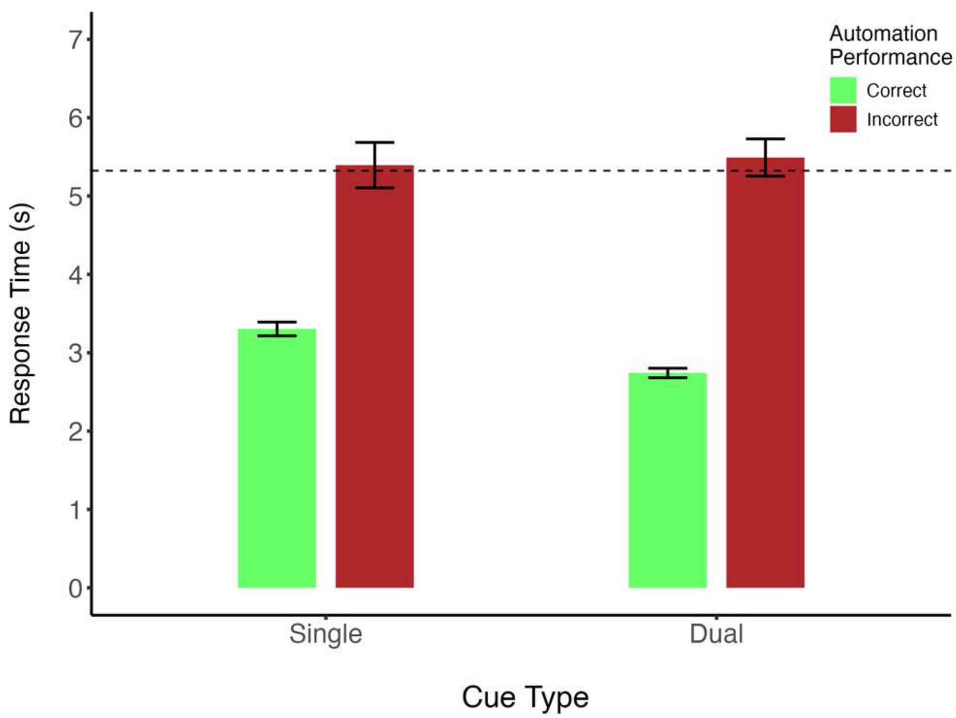

Figure 10. Experiment 1: The mean response time for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

A 2 (cue type) x 2 (automation performance) repeated-measures ANOVA was conducted to examine the automation bias for response time as a function of cue type (single versus dual) and automation performance (correct versus incorrect). Results revealed no significant main effect of dual cueing, $F(1, 46) = 1.23$, $p = .274$, $\eta_p^2 = 0.03$. There was a significant main effect of automation performance, $F(1, 46) = 118.60$, $p < .001$, $\eta_p^2 = 0.72$, indicating that response time was faster when automation was correct ($M = 3.11$ s) than when automation was incorrect ($M = 5.45$ s). The interaction between cue type and automation performance was not statistically significant, $F(1, 46) = 2.91$, $p = .09$, $\eta_p^2 = 0.06$. While the interaction was not statistically significant, it should be noted that dual cues did result in faster response times than single cues when the automation was correct ($t(46) = 5.18$, $p < .001$, 95% CIs [0.35, 0.79], $d = 0.51$.

Percent Error. The same exact analyses were conducted for percent error. The mean percent error data are shown in Figure 11.

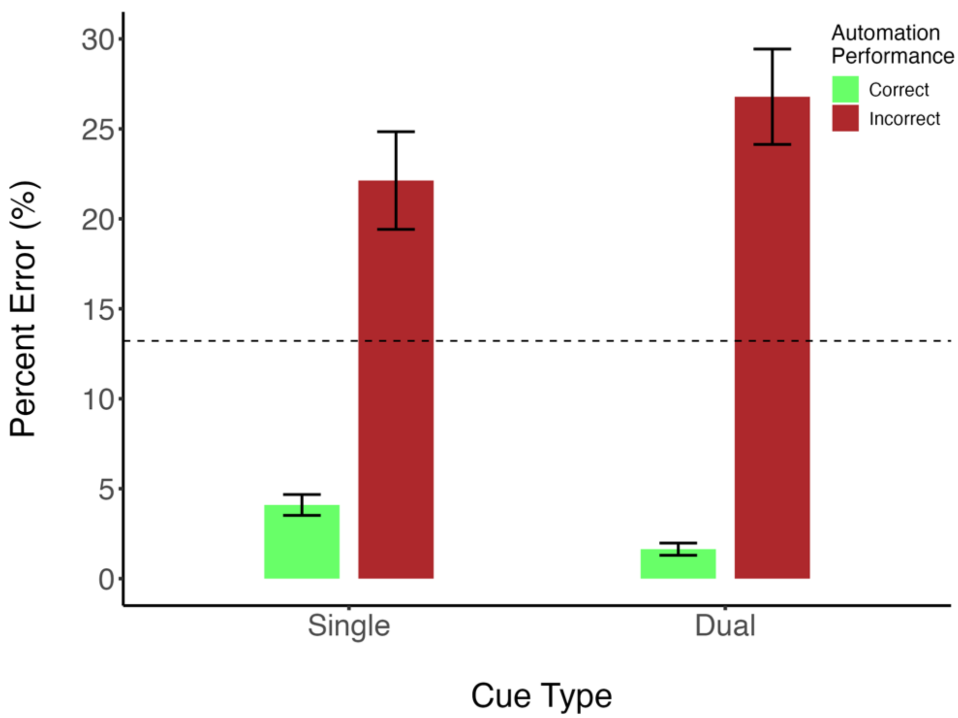

Figure 11. Experiment 1: The mean percent error for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

The main effect of dual cueing was not statistically significant, $F(1, 46) = 0.44$, $p = .51$, $\eta_p^2 < .01$. The main effect of automation performance was significant, $F(1, 46) = 55.08$, $p < .001$, $\eta_p^2 = 0.54$, revealing that percent error was much greater when automation was incorrect ($M = 24.66\%$) than when correct ($M = 2.76\%$), and greater than the unaided condition ($M = 13.21\%$) indicated by the dashed line in Figure 11. Critically, there was a significant interaction between cue type and automation performance, $F(1, 46) = 4.51$, p = .04, $\eta_p^2 = 0.09$. When automation was correct, dual cues ($M = 1.6\%$) decreased errors more than single cues ($M = 4.0\%$), ($t(46) = 3.22$, $p = .002$, 95% CIs [0.01, 0.04], $d = 0.70$). However, when automation erred, there were no differences in error rates, $t(46) = -1.41$, $p = .16$, 95% CIs [-0.12, 0.02], $d = 0.20$, although trending toward greater error in the dual cue condition. As a reminder, there is a low number of trials for which the automation can error, meaning that the analysis has lower statistical power.

In summary, participants show an overall automation bias relative to the unaided search. Interestingly, prior work has shown that the cue that is most effective results in a greater cost when it errs (Warden et al., 2023). However, in this case, the bias was similar regardless of whether the searcher was using a single or dual cueing aid, suggesting no greater cost of dual cues (the more effective cues) compared to single cues.

Effect of Cue Precision and Number of Cues: Redundancy Gain

As previously mentioned, cue conditions were also categorized based on their level of precision in terms of the type of information they conveyed. Spatial cues conveyed either global spatial information, local spatial information, or a combination of global and local spatial information. The global, local, and global combined with local cues were used to examine the effect of redundancy gains on performance. Redundancy gains were defined to occur when the combination of a global and local cue improves performance beyond that of either cue independently. Data were collapsed across both levels of cue reliability, and the no-cue condition was excluded from the analysis. Additionally, cues that conveyed identity information were excluded from the following analysis.

Response Time. The mean response time across cue types is presented in Figure 12.

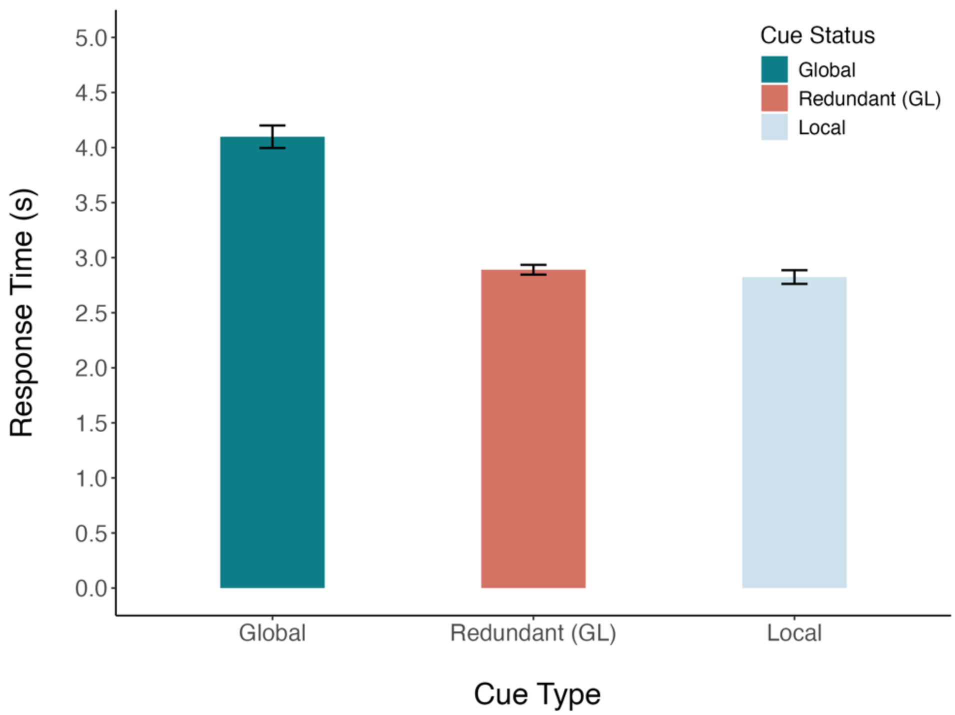

Figure 12. Experiment 1: The mean response time for cue type. Error bars represent one standard error of the mean.

A one-way repeated-measure ANOVA revealed a significant effect of cue type on response time, $F(1.41, 64.98) = 81.59$, $p < .001$, $\eta_p^2 = 0.64$. Pairwise comparisons (paired *t*-test) showed that response time was significantly slower for the global cues ($M = 4.10$ s) compared to the local ($M = 2.82$ s; $t(46) = -9.77$, $p < .001$, 95% CIs [-1.55, -1.02], $d = 1.09$) and redundant cues ($M = 2.89$ s; $t(46) = 9.45$, $p < .001$, 95% CIs [0.96, 1.47], $d = 1.02$). More importantly, there was no significant response time difference between local and redundant cues, $t(46) = -1.03$, $p = .307$, 95% CIs [-0.21, 0.07], $d = 0.07$. This finding indicates there was no redundancy gain, meaning there was no additional search benefit of global cues.

Percent Error. The mean percent error across cue types is presented in Figure 13. Results revealed a significant effect of cue type on percent error, $F(1.36, 62.63) = 23.04$, $p < .001$, $\eta_p^2 = 0.33$. Pairwise comparisons (paired *t*-test) showed the same pattern of results as the response time data. Specifically, percent error was significantly higher for global cues ($M = 6.37\%$) compared to the local ($M = 2.09\%$; $t(46) = -5.22$, p < .001, 95% CIs [-0.06, -0.03], $d = 0.84$) and redundant cues ($M = 2.57\%$; $t(46) = 4.88$, $p < .001$, 95% CIs [0.023, 0.055], $d = 0.71$). Like response time, there was no significant difference between local and redundant cues, $t(46) = -1.22$, $p = .228$, 95% CIs [-0.01, 0.003], $d = 0.19$.

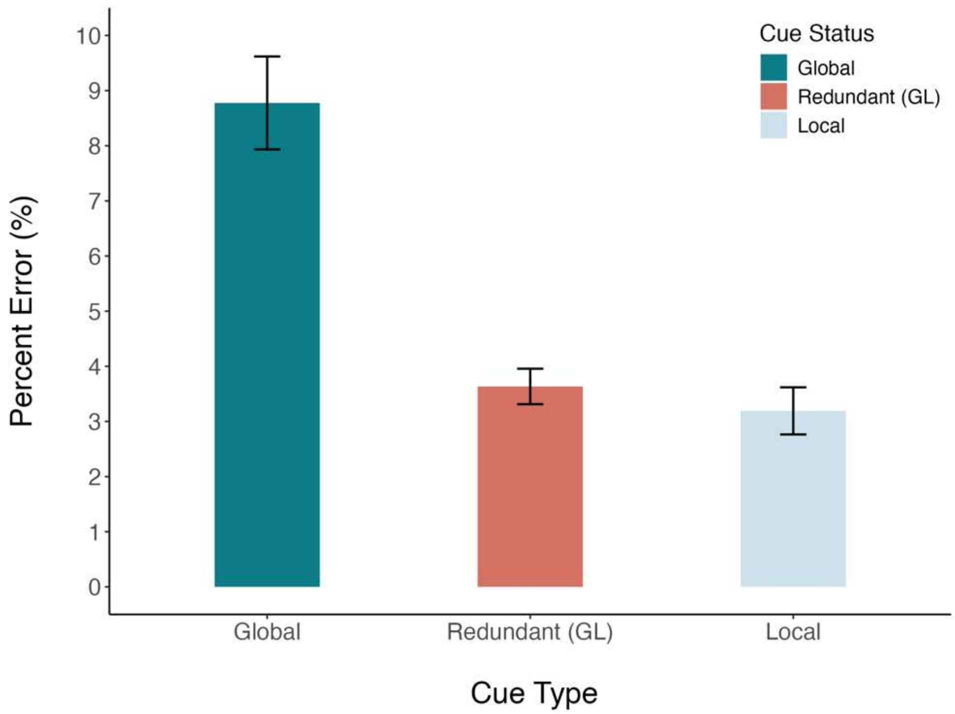

Figure 13. Experiment 1: The mean percent error for cue type. Error bars represent one standard error of the mean.

Collectively, the pattern of results for response time and percent error show that both local and redundant (global + local) cues improve overall search performance over global cues: there was no redundancy gain. In the context here, the local spatial information is sufficient to facilitate an efficient search, and the addition of global information in the redundant cue does not provide an additional advantage.

Effect of Attentional Tunneling

The effect of attentional tunneling requires examining performance when a cue unexpectedly fails to cue a high priority object. This could happen if the automation was not designed to detect a certain kind of hazard in the environment. For example, the sensors in an automated vehicle might be designed to cue attention to pedestrians, but not to cue attention to other hazards like a large pothole in the road. In such "black swan" (unexpected) or "gray swan" (very low expectancy) instances (see Wickens et al., 2009) it is important to see whether cues that direct attention to certain kinds of targets result in an automation-based attention tunneling, where the user narrows their attention on where the cue is directing them, resulting in missing other uncued but even more important information present in their visual field.

The magnitude of attention tunneling was examined using response time and percent error data. Only trials where a high priority target appeared were included. For response time data, trials where the high priority target was not selected were excluded because these trials had no associated response time. The analysis examined whether the cue status of the routine target, either cued or uncued, impacted the searcher's ability to find the uncued, less expected high priority target. All data were collapsed across cue reliability, cue type, and set size.

A 2 (target type) x 2 (cue status) repeated-measures ANOVA was conducted to examine the effect of target type (high priority versus routine target) and routine target cue status (cue, uncued) on both response time and percent error. The data from one participant was removed because they did not select any high priority targets.

Response Time. The response time for target type as a function of cue status are shown in Figures 14.

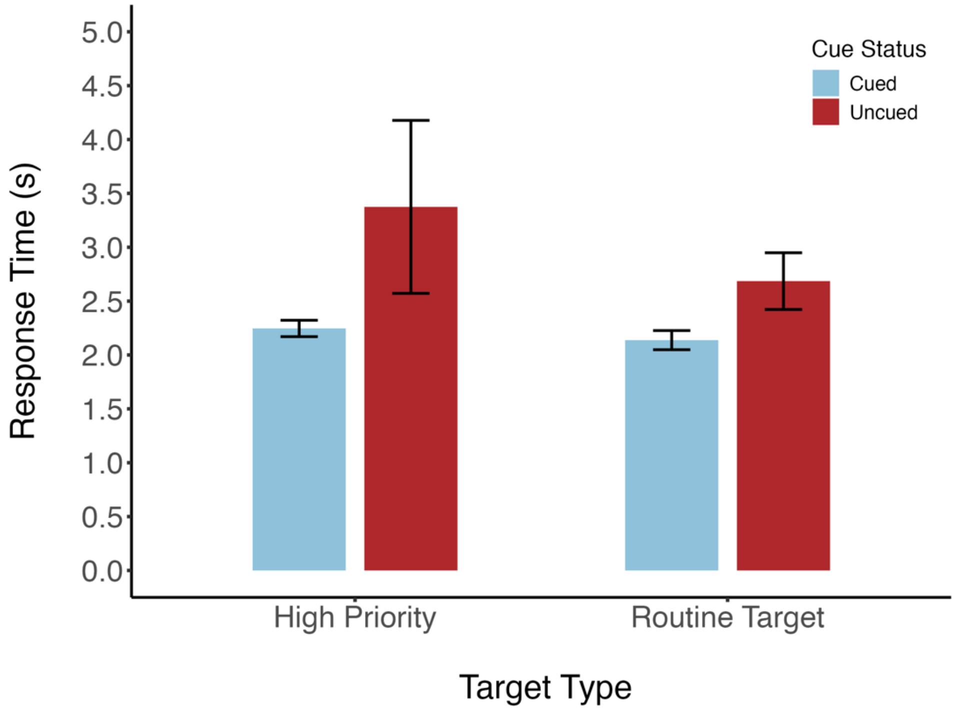

Figure 14. Experiment 1: The mean response time for target type (high priority and routine) as a function of whether the routine target was cued or uncued. Error bars represent one standard error of the mean.

A total of seven participants were removed from the analysis because they failed to find either object on uncued trials. The ANOVA revealed that response times for both the routine and high priority target were significantly faster when the routine target was cued (M = 2.17 s) than uncued (M = 2.91 s), $F(1, 34) = 6.89$, $p = .013$, $\eta_p^2 = 0.17$. Neither the effect of target type ($F(1, 34) = 0.297$, $p = .589$) nor the interaction ($F(1, 34) = 0.561$, $p = .459$) were significant.

Percent Error. The percent error or miss rate for target type as a function of cue status is shown in Figure 15. The ANOVA revealed a significant main effect of target type (routine, high priority), $F(1, 45) = 55.88$, $p < .001$, $\eta_p^2 = 0.55$, indicating that percent error is greater for the high priority target (M = 45.58%) compared to the routine target (M = 10.42%). While the main effect of cue status (cued, uncued) was not significant, $F(1, 45) = 0.64$, $p = .427$, $\eta_p^2 = 0.01$, there was a significant interaction between target type and cue status, $F(1, 45) = 5.77$, $p = .02$, $\eta_p^2 = 0.11$. Pairwise comparisons (paired *t*-test) show that cued routine targets *(M* = 5.62%) led to significantly lower errors than uncued routine targets (M = 15.22%), reinforcing the overall effect of cueing. However, for the less expected high priority target, there were no significant differences in error rate as a function of whether they were cued (*M* = 47.67%) or not (*M* = 43.48%). However, there is a trend suggesting greater errors for the high priority when the routine target was cued.

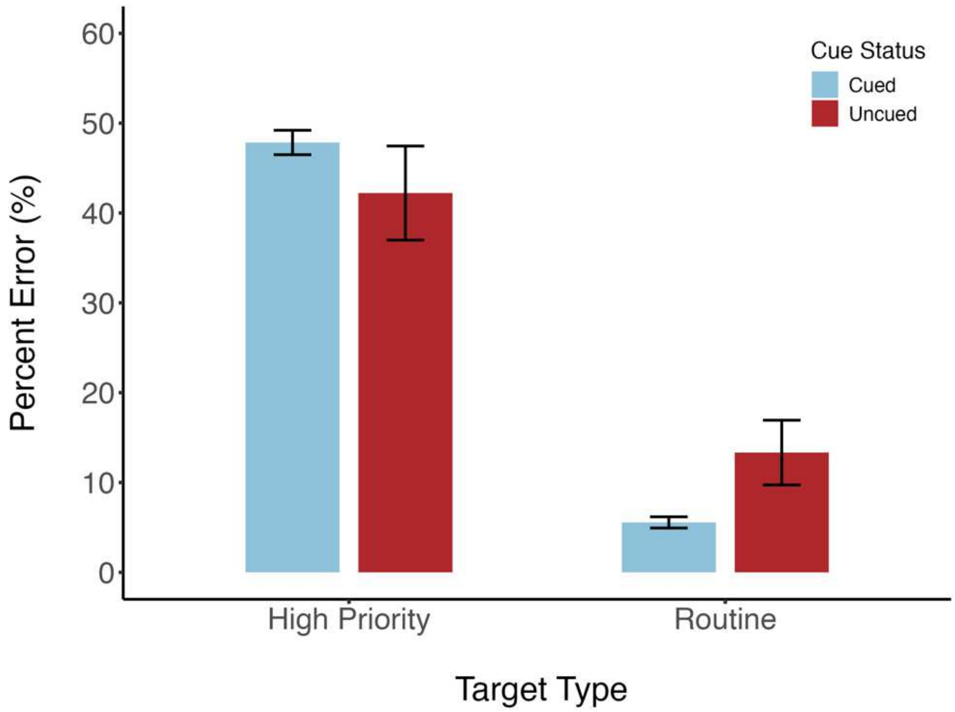

Figure 15. Experiment 1: The mean percent error for target type (high priority and routine) as a function of whether the routine target was cued or uncued. Error bars represent one standard error of the mean.

To further examine the effects of attention tunneling and whether dual cueing induces greater attention tunneling than single cueing, the data were categorized based on target type (high priority, routine) and cue status (no cue, single cue, dual cue).

Response Time. Figure 16 shows mean response times. The ANOVA showed that response time for routine targets (*M* = 2.38 s) was significantly faster than high priority targets (*M* = 2.52 s), $F(1.57, a53.43) = 5.62$, $p = .010$, $\eta_p^2 = 0.14$. Neither the effect of target type ($F(1, 34) = 0.194$, $p = .662$) nor the interaction ($F(1.24, 42.10) = 0.676$, $p = .446$) were significant.

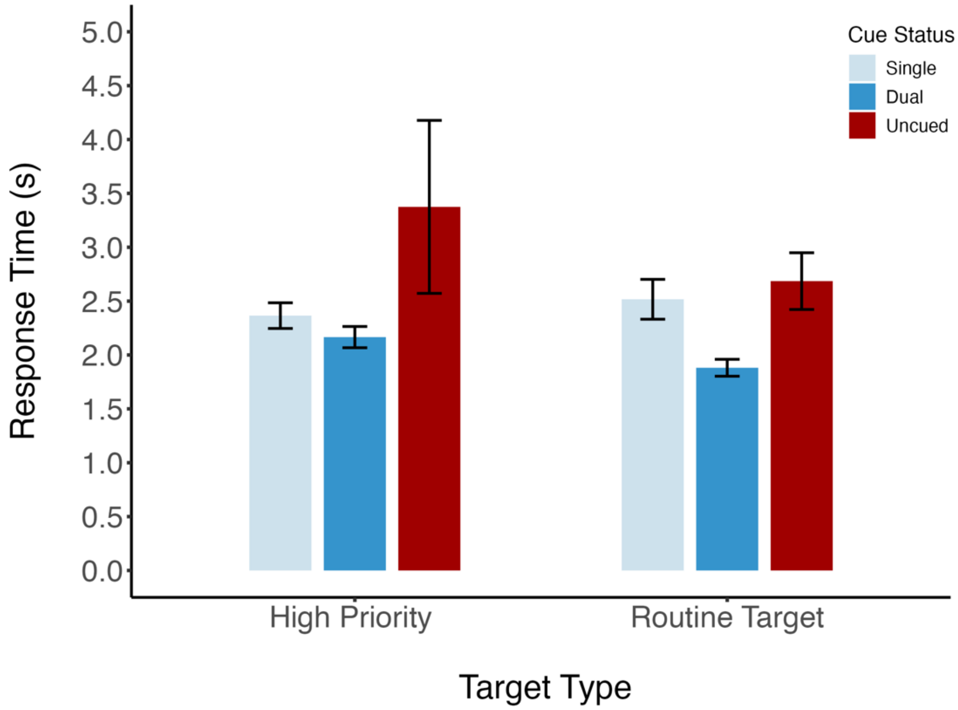

Figure 16. Experiment 1: The mean response time for target type (high priority and routine) as a function of cue status (dual, single, uncued). Error bars represent one standard error of the mean.

Percent Error. Figure 17 shows mean percent errors. The ANOVA revealed a significant main effect of target type, $F(1, 45) = 77.54$, $p < .001$, $\eta_p^2 = 0.63$, indicating that percent error was greater overall for the less expected high priority target ($M = 46.26\%$) compared to the routine target ($M = 8.82\%$). The main effect of cue status was not significant, $F(1.27, 57.24) = 0.58$, $p = .488$, $\eta_p^2 = 0.01$. There was a significant interaction between target type and cue status, $F(1.34, 60.19) = 4.84$, $p = .022$, $\eta_p^2 = 0.10$. The pattern of this interaction is the same as in Figure 14. However, here it also signals that the attentional tunneling cost of cueing the routine target is identical for dual and single cues, even as the former had shown benefits across all trials (e.g., Figure 8)

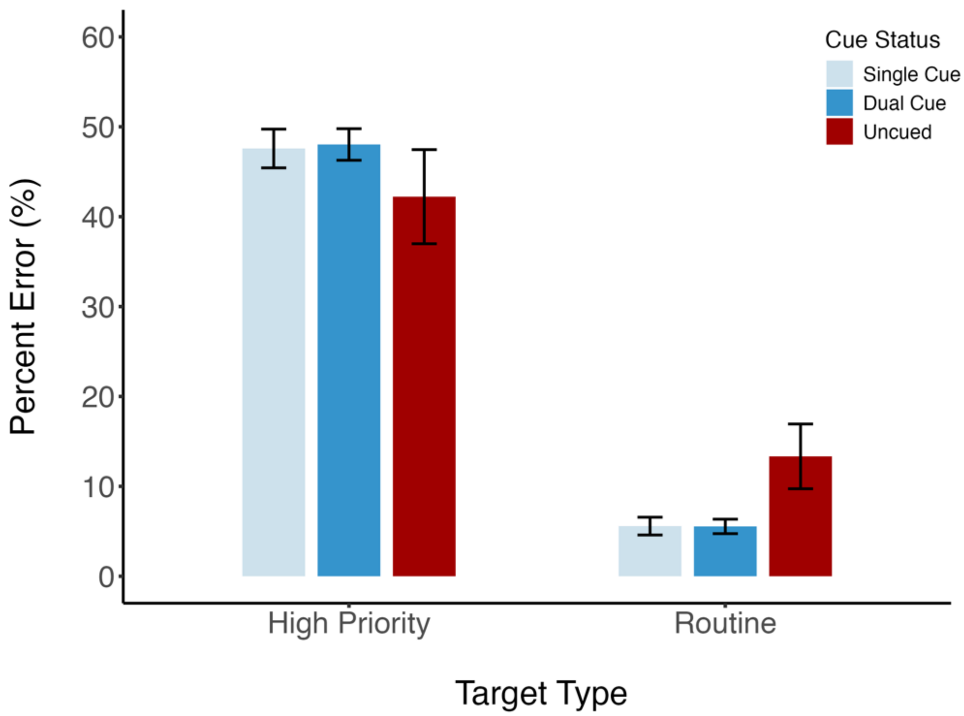

Figure 17. Experiment 1: The mean percent error for target type (high priority and routine) as a function of cue status (dual, single, uncued). Error bars represent one standard error of the mean.

Prior work found that the cue that helped performance the most (i.e., the arrow cue), hurt performances the most when it was wrong (Warden et al., 2023). Thus, the following analysis seeks to examine whether the "better" cue (here, the gaze guidance line) increases attentional tunneling more than the "worst" cue (here, the global arrow) for an uncued, less expected high priority target. The same analysis for response time and percent error above was applied to these specific cue types to examine whether the attentional tunneling is more pronounced in the "better cue" case. The data for this examination are shown in Figures 18 and 19.

Response Time. Figure 18 shows mean response time data. The ANOVA revealed a significant effect of cue status, $F(1.55, 32.49) = 6.58$, $p = .007$, $\eta_p^2 = 0.24$, showing that the gaze guidance line (M = 1.77 s) led to the fastest response time and the uncued condition time (M = 2.91 s) led to the slowest response. Neither the effect of target type ($F(1, 21) = 1.74$, $p = .202$) nor the interaction ($F(1.66, 34.85) = 1.03$, $p = .356$) was significant.

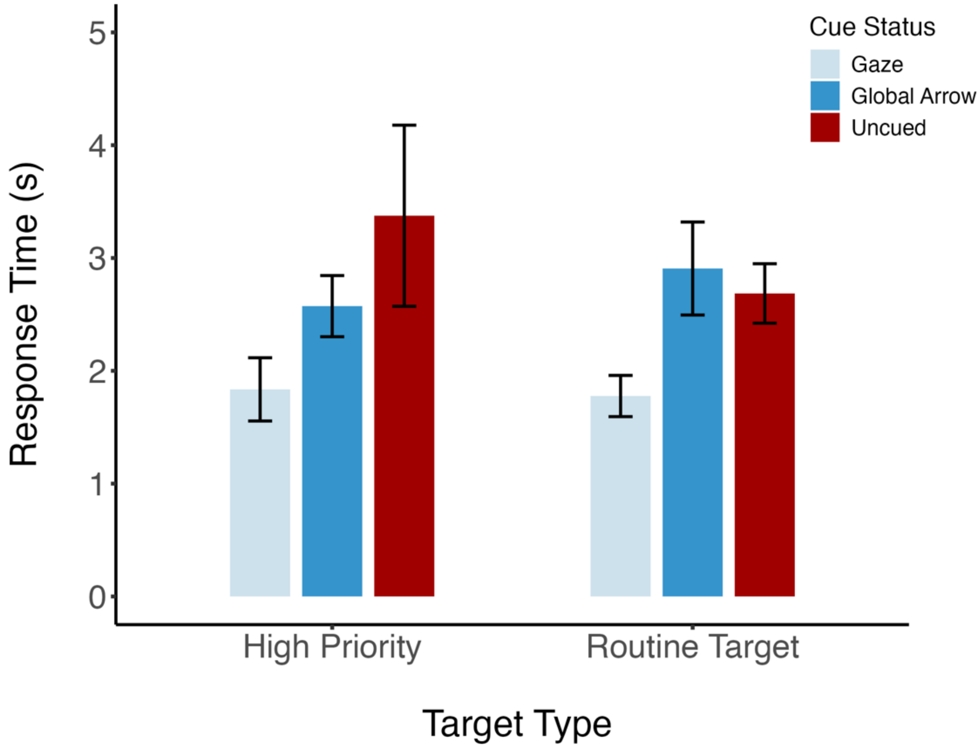

Figure 18. Experiment 1: The mean response time for target type (high priority and routine) as a function of cue status (gaze, global arrow, uncued). Error bars represent one standard error of the mean.

Percent Error. The data for precent error are shown in Figure 19. Results revealed a significant main effect of target type, $F(1, 45) = 64.36$, $p < .001$, $\eta_p^2 = 0.58$. The main effect of cue type was not statistically significant, $F(2.85, 128.6) = 0.36$, $p = .78$, $\eta_p^2 = 0.008$. There was a significant interaction between target and cue type, $F(2.71, 122.02) = 4.254$, $p = .007$, $\eta_p^2 = 0.09$. This interaction revealed that the gaze guidance cue, which directly linked to the routine target, provided the most accurate performance for the routine target, but there was a marginally significant difference in percent error for the high priority target when the routine target was cued with the gaze guidance cue compared to when the routine target was uncued, $t(45) = 1.88$, $p = .067$, 95% CI [-0.01, 0.23], $d = 0.27$. This finding hints that the gaze guidance cue may cause attentional tunneling for the less expected high priority target appearing at the same time as the routine target.

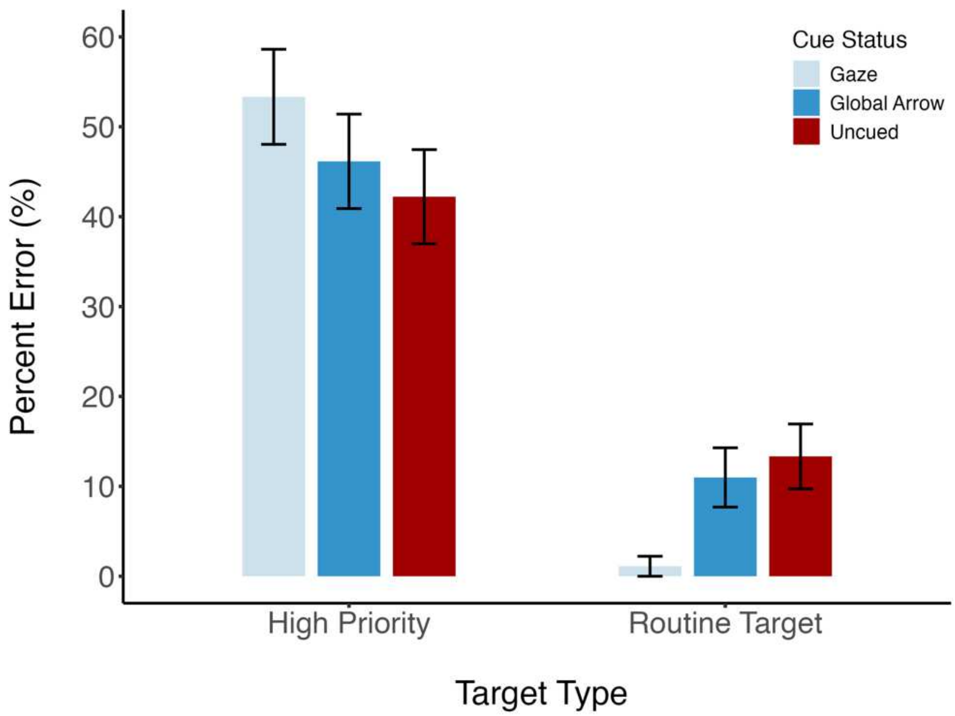

Figure 19. Experiment 1: The mean percent error for target type (high priority and routine) as a function of cue type (gaze, global arrow, uncued). Error bars represent one standard error of the mean.

In summary, evidence of attentional tunneling was only weakly observed in percent error data. Overall, there was a trend suggestive of attentional tunneling for locating the high priority target when the routine target was cued. Additionally, results suggest that a cued routine target with the gaze guidance line marginally increased the error rate for a less expected high priority target. In general, dual cueing did not further enhance attentional tunneling more than single cues. Response time data showed a decrease in search time for the less expected high priority target when the routine target was cued rather than the predicted increase in search time; this can be attributed to the high expectancy of the routine target compared to the low expectancy of the high priority target.

Discussion

The current experiment examined the influence of cueing generally, and specifically, dual-cueing for a static visual search task conducted on a wide-angle desktop display. Replicating prior work (Warden et al., 2022; Warden et al., 2023), the current findings support H_1, showing an overall performance benefit of cueing compared to an unaided search. In general, cueing improves search time and accuracy, and cues become more beneficial when there are more items in the scene to search through, highlighting their role in attention guidance during complex visual search tasks. Additionally, H_2 was confirmed, showing that dual cues enhanced search performance and accuracy more than single cues when cues were perfectly reliable, suggesting that dual cues configured in a global-local combination effectively manage response time and accuracy, especially as the complexity and size of the search scene increase.

However, cueing effectiveness is hurt by imperfectly reliable automation. Confirming H_{1A}, performance degraded when the cueing aids erred compared to when they were perfect, suggesting that people exhibited an automation bias for cues in general. Disconfirming H_{2A}, there was no significant difference between dual and single cues when the automation was incorrect. Overall, these findings highlight the degrading effect imperfect automation has on performance with decision aids like attentional cues. Additionally, dual cues did not significantly enhance the negative effects of imperfect automation relative to single cues, which may speak favorably to dual cueing systems that may err on occasion. However, it should be noted that there was a non-significant trend in percent error data suggesting that dual cues may result in a greater automation bias than single cues when the automation was incorrect. It should be reiterated that the number of trials for which the automation could error was low, resulting in low statistical power. With sufficient power, it may be the case that the increased complexity of managing two simultaneous cues intensifies the negative effects of imperfect automation.

There is some evidence from the response time data that participants did not always blindly follow the automation. As seen in the response time data, there is an increase in search time for both the cued and uncued conditions as the set size increased, suggesting that participants partially resisted blindly following the automation. The slowed search time as the set size increased in the cued condition suggests that some checking of the other items in the scene occurred. However, this additional checking may have been linked to also looking for the less expected high priority target, rather than confirming the target was the correct one.

Despite prior research suggesting that cueing causes attentional tunneling (Yeh et al., 1999), the effect was not robustly supported in the current experiment. Collectively, evidence of attentional tunneling was subtle and only weakly observed in percent error data, contradicting H_{1B} and H_{2B}. But the number of trials for which a less expected high priority target appeared was limited, resulting in low statistical power.

Despite the lack of significance, a consistent pattern emerged, indicating that cueing the routine target increased error rate for the less expected high priority targets. This was marginally significant for the gaze guidance line, suggesting the best cue led to greater attentional tunneling. But dual cues did not exacerbate attentional tunneling compared to single cues, suggesting that additional cues did not further distract from a less expected high priority target.

The response time data revealed an opposite pattern of results, specifically that cueing the routine target led to faster identification of the less expected high priority target. This effect likely reflects the fact that participants were instructed to prioritize the high priority target, and, therefore, they developed a search pattern that benefited response times for that target. The results may also indicate a speed-accuracy tradeoff, where they attempted to rapidly locate the less expected high priority target at the cost of accuracy.

Partially supporting H_3, the results show that performance improved with both the local and redundant (global+local) cues compared to the global cue alone and that performance was equivalent between the local and the redundant cue. This finding reinforces that more precise local cues in the current context are just as effective or more effective than cues that also provide global spatial information. Overall, redundant cueing improves search performance compared to a global cue alone, but the advantage of a redundant cue over a local cue may depend more on the complexity of the search environment. While there was an overall main effect of dual cueing, such that dual cues led to better performance, the current experiment suggests this is driven by the local cue alone. This can be attributed to the fact that participants could exploit peripheral vision, where they could see the local cue in their periphery. It may be the case that there is an advantage of the redundant cue compared to both global and local cues in a larger search scene when part of the search field is invisible, and hence, peripheral vision cannot be exploited. This possibility is investigated in Experiments 2 and 3.

EXPERIMENT 2: STATIC VISUAL SEARCH USING AN AR-HMD

Description

The primary purpose of Experiment 2 is to examine cue effectiveness when cues are presented with an augmented reality HMD (AR-HMD) during a large, 157-degree realistic, and static search task. Another primary purpose of this experiment is to assess whether the frame-of-reference (i.e., world versus screen referenced coordinates) of AR-HMD cues impacts search performance and to examine whether findings from the 2D desktop monitor generalize to an AR-HMD.

Method

Participants. A total of 38 participants in an introductory psychology course at Colorado State University received course credit in exchange for completing the experiment. No participants from Experiment 1 participated in Experiment 2. All participants had self-reported normal or corrected-to-normal vision and were screened for colorblindness using an electronic version of Ishihara's test. No demographic information was collected.

Stimuli and Apparatus. Participants completed the experiment using a Microsoft HoloLens 2 (HL2), an augmented reality (AR) optical see-through head-mounted display that embeds and overlays virtual content onto the real-world. The HL2 has a field of view of 43 degrees in the lateral direction and 29 degrees in the vertical direction. The cue conditions were created in the game engine Unity (version 2020.3.28f1) and used the XR interaction toolkit. The unit of measure in Unity was 1 unity measure, which was 40.5 inches in the present experiment. The search scene consisted of the same 2D scene images used in Experiment 1, except that only 4 unique scenes were used to project a 238-inch by 120-inch virtual image in the real-world environment. The virtual background images were projected approximately 24 inches away from the participant using the HL2, and the participant was positioned at the midpoint of the scene.

The search scene consisted of the same 128 routine objects from Experiment 1 plus 64 new routine objects (of the same categories as Experiment 1) for a total of 192 unique routine objects that were uniformly distributed (see Figure 20 below). Given the dimensions of the virtual scene, some objects inherently fall outside of the FOV of the device, requiring participants to look around for the objects in the scene.

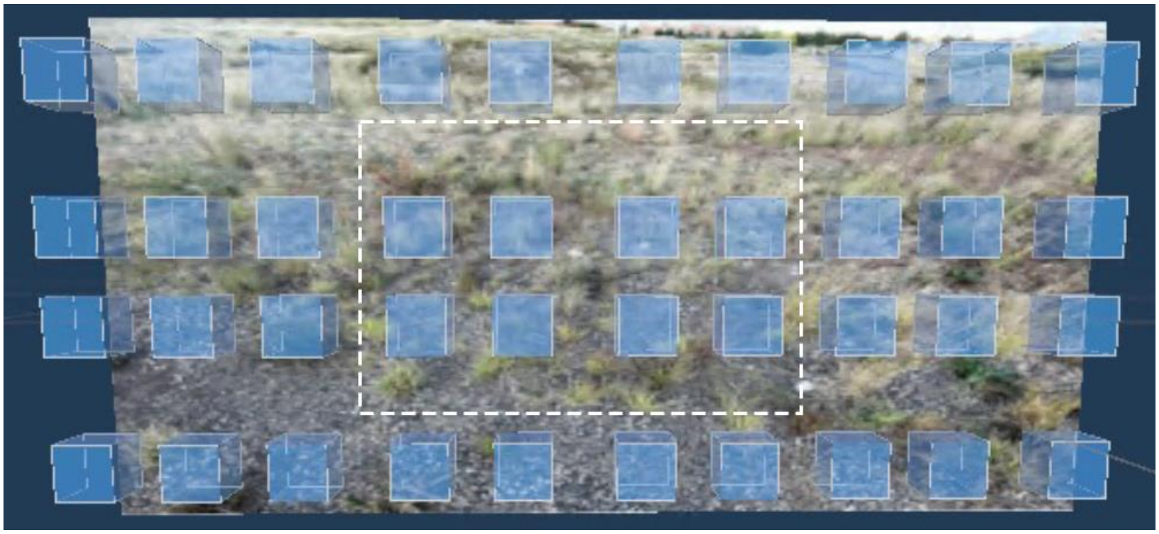

Figure 20. An image of the virtual scene with all possible object locations uniformly distributed across the scene. The cubes represent the possible object locations where routine and high priority objects could be positioned. The white dashed rectangle represents the FOV of the HMD. Objects inside this boundary were inside of the searcher's immediate FOV. Objects outside were out of the searcher's immediate FOV and require a head movement to bring them into view.

The scene also consisted of an uncued, less expected high priority object. The same high priority object images were used as Experiment 1, except that only four or the original five were used. The bomb image was removed for two reasons: (1) to minimize memory load and (2) the type of bomb differed from the others; therefore, removing it ensured consistency across the less expected high priority objects. As in Experiment 1, objects in the foreground, middle ground, and background were scaled to simulate depth in the search field. All objects were located approximately 24.3 inches (i.e., 0.6 unity measures) apart. Object images were post-processed the same way as Experiment 1.

The same 16 cue conditions from Experiment 1 were used in Experiment 2 (see Figure 4). Unlike Experiment 1, cues now varied based on their display imagery, which could be in either world-reference coordinates (conformal imagery) or screen-referenced coordinates (nonconformal imagery). As a reminder, world-referenced cues (gaze guidance line, global arrow, local arrow, highlight) are those that have a one-to-one correspondence with the (x, y, z) location of an object in the real-world and depend on the orientation of the head, meaning that the cue always points to, or directly overlays the object in the real-world as the head rotates. All cueing aids only cued the routine target objects. Screen-referenced cues (minimap, icon) are cues that correspond to the (x, y) coordinates on the display of the device and do not depend on the orientation of the head. Figure 21 below shows examples of the single cue conditions and Figure 22 shows examples of the dual cue conditions viewed in the AR-HMD

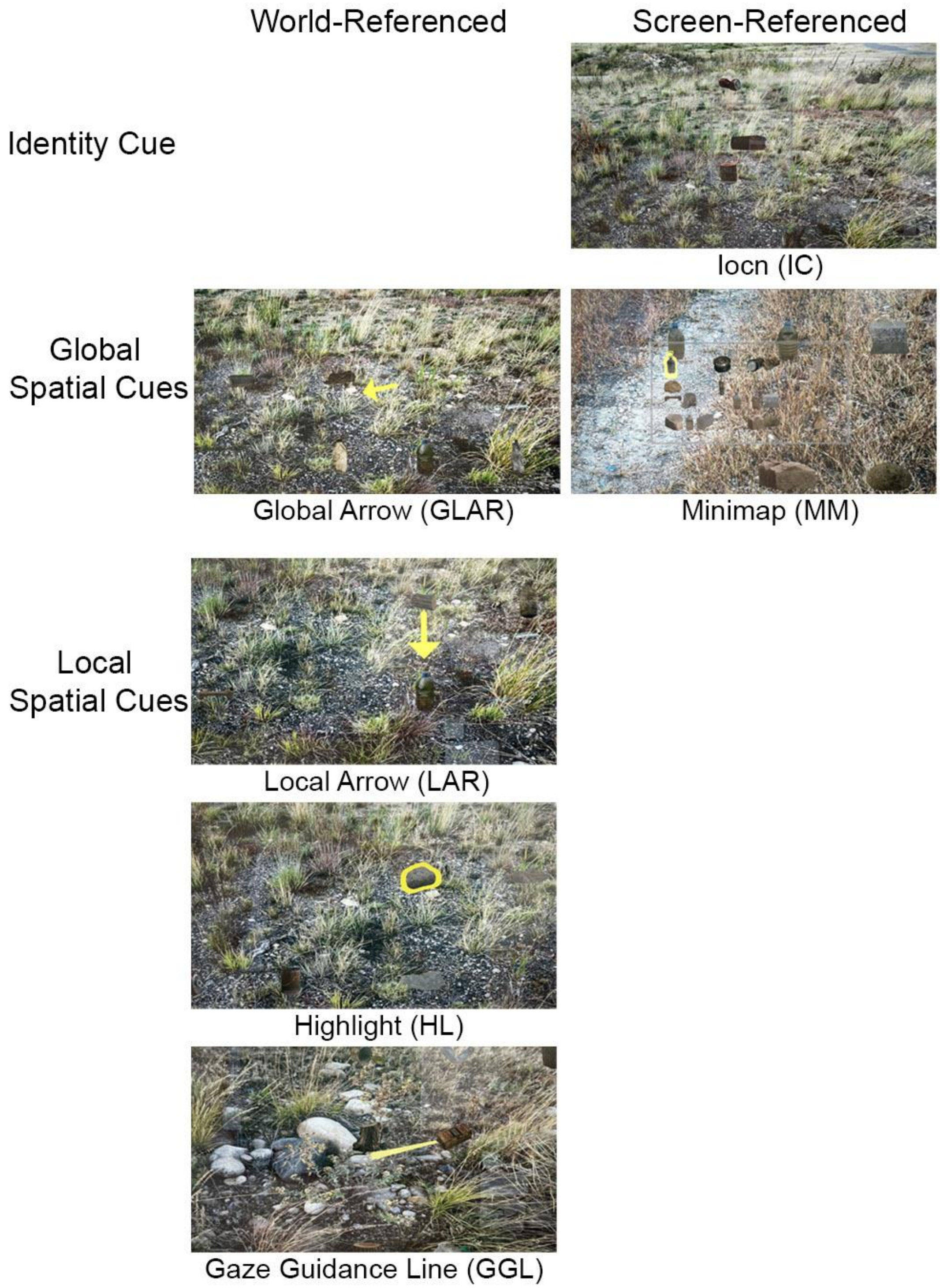

Figure 21. Screen captures from the experiment presented in the AR-HMD showing each of the single cue conditions, their corresponding content (i.e., identity, global or local spatial information), and the type of display imagery used (world or screen referenced). Note that the minimap and icon cue consisted of a transparent outline to distinguish it from the search scene.

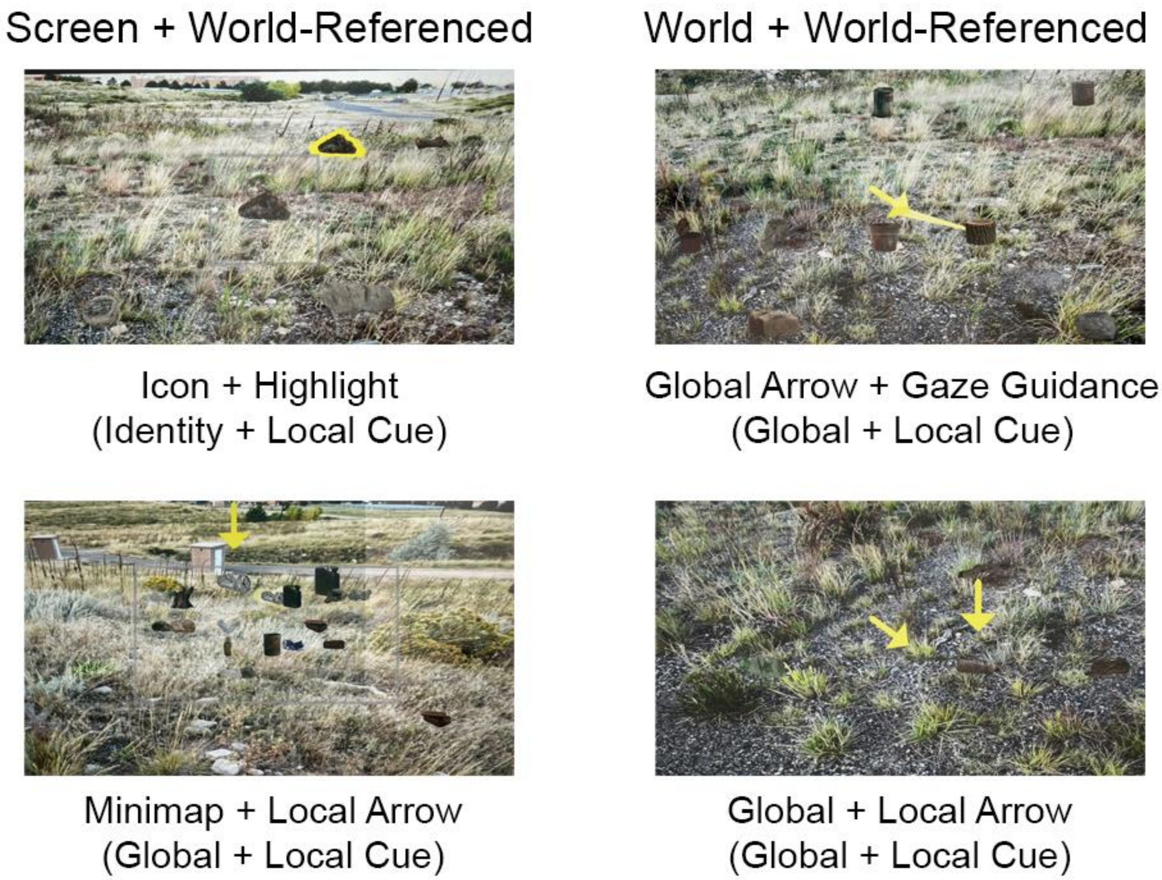

Figure 22. Screen captures from the experiment presented in the AR-HMD showing an example of the dual-cues. Dual-cues contained either screen + world-referenced or world + world referenced cues. The content conveyed by the cue is also listed below each cue type. Note that the minimap and icon cue consisted of a transparent outline to distinguish it from the search scene.

Like Experiment 1, these cueing aids were either perfect (i.e., 100% correct) or imperfect (i.e., 83% correct). For the imperfect trials, cues errored 17% of the time by cueing the wrong target object, representing a false alarm. When the cue errored by pointing to the wrong target, the correct target was always present and located 202.5 inches away from the wrongly cued target. The less expected high priority object was never cued and only appeared 17% of the time for each cue condition, including the no cue. On trials where the less expected high priority object appeared, it was located, on average, a maximum of 142 inches away from the routine target for that trial. This resulted in approximately three objects between the routine and high priority targets on a given trial. Figure 23 below shows an example of a high priority target in the search scene.

Figure 23. An example of the search scene from the user's perspective showing the uncued, less expected high priority object and the constrained field-of-view created by the device.

Experimental Design. The experiment was a 16 (cue type) x 2 (cue reliability) within-subject design. Cue type could be categorized in various ways depending on the type of analysis conducted. These different categorizations are described in the Results section. The set size for all searches was 18 to ensure participants could complete the task within 1 hour.

Before each trial, the AR-HMD displayed text indicating the cue condition for 1.5 s, followed by a 300 ms fixation cross, and then the target for that trial appeared on the HMD display to study for 3 seconds. At the beginning of each trial, participants were required to look forward to ensure a default starting position for each trial. Once the trial began, participants scanned the virtual scene for the routine target. As participants scan the search scene, a head position-sensitive algorithm based on the camera sensors renders a hitbox (i.e., transparent cube outline) around the object they have oriented their heads toward. This hitbox signals to the participant which object can be selected.

Participants completed two blocks, which were counterbalanced by cue reliability (see Experiment 1 for details). Participants completed all cue conditions, including the no cue, for each reliability block. The cue condition sub-block was counterbalanced within the cue reliability blocks. Within each cue condition sub-block, there was a total of 6 randomized trials for each unique cue condition. The high priority object will appear on 1 of the 6 trials (approximately 17% of the time). For the imperfect cue condition, the cue errored on 1 of the 6 trials (approximately 17% error rate). When the cue errors, it will identify an incorrect target. Each reliability block consisted of 96 trials, for a total of 192 trials in the experiment. Participants also completed 1 practice trial for each cue condition for a total of 16 practice trials before they started the experiment. Participants received text feedback (i.e., “correct” or “incorrect”) on the practice trials.

Task. Participants completed a 158-degree static visual search task using an AR-HMD. Participants sat in a stationary chair positioned so that they were facing the middle of the virtual scene and approximately 24 inches away from the scene. Like Experiment 1, they were instructed to find the routine target and told a less expected high priority object would appear occasionally and that this object should take precedence. The instructions for the task were the same as Experiment 1 except that the response input device and key presses changed (see Procedure below for details; see Appendix B). After the search task, participants completed the same trust in automation survey and answered the same questions about each cue condition used in Experiment 1 (see Appendix D and E).

Procedure. Participants gave consent to participate after reading and signing the consent documentation before starting the experiment. Next, participants completed an electronic colorblindness test to screen for red-green colorblindness deficiencies. Then, they read the experimental instructions and were shown example images of routine objects and the four high priority objects with low expectancy. Participants calibrated their eyes to the HL2 device and then sat in a stationary chair in front of the virtual search scene to ensure an approximate visual angle of 158 degrees. Before the experimental trials, they completed the practice trials. Participants used a wireless Logitech Bluetooth keyboard to select objects during the search task. They pressed the ‘*right shift’* button to select the routine target and the high priority target, followed by the ‘*left control’* button to continue to the next trial. The entire experiment lasted approximately 1 hour.

Results

The experimental design was the same as Experiment 1 except that only one set size (n = 18) was used to ensure participants could complete the entire experiment within 1 hour. Before conducting any analyses, the data were examined for outliers in the same way as in Experiment 1. Based on these criteria, three participants were deemed outliers and removed from the analysis. Additionally, two participants only completed half of the experiment and were removed due to incomplete data. A total of five participants were excluded from the analysis. All remaining data (N = 33) was analyzed using R Studio.

The assumption of normality was tested using the Shapiro-Wilk normality test, and normality was violated for both response time and error data despite data transformations (*ps* < .05). Therefore, all reported ANOVA analyses include the Greenhouse-Geisser (GG) correction to account for such violations.

Like Experiment 1, all response time data includes correct and incorrect responses. When analyzing the response time data for routine targets, trials in which participants did not make a response or only selected the high priority target were removed from the analysis to ensure that response time pertained only to the routine targets that were selected. Percent error data for the routine targets included all trials.

Descriptive Statistics for Perfect Cue Conditions

Table 3 and 4 below present cue response times and percent errors for the perfect cueing condition only. The cue conditions included are icon (IC), minimap (MM), global arrow (GLAR), local arrow (LAR), highlight (HL), gaze guidance line (GGL), and the no-cue condition.

Table 3. Experiment 2: Mean response times, measures in seconds, for each single cue and the respective dual-cue combinations. Standard deviations are presented in parentheses.

	Single Cues	**IC**	**MM**	**GLAR**	**LAR**	**HL**	**GGL**
IC	**8.85 *s*** (*7.19*)				**5.11 *s*** (4.04)	**5.38 *s*** (4.47)	**3.14 *s*** (3.73)
MM	**7.37 *s*** (6.37)				**4.34 *s*** (2.91)	**4.68 s** (4.27)	**3.27 *s*** (2.90)
GLAR	**4.44 *s*** (3.68)				**3.05 *s*** (2.18)	**3.04*s*** (3.04)	**2.75 *s*** (2.58)
LAR	**5.68 *s*** (5.47)						
HL	**4.98 *s*** (3.21)						
GGL	**2.88 *s*** (2.45)						
No-Cue	**9.11 *s*** (8.12)						

Table 4. Experiment 2: Mean percent errors for each single cue and the respective dual-cue combinations. Standard deviations are presented in parentheses.

	Single Cues	**IC**	**MM**	**GLAR**	**LAR**	**HL**	**GGL**
IC	**16.67%** (*37.37*)				**3.03%** (*17.19*)	**4.17%** (*20.03*)	**1.52%** (*12.25*)
MM	**13.02%** (*33.74*)				**2.96%** (*16.98*)	**1.52%** (*12.25*)	**0.52%** (7.22)
GLAR	**4.41%** (*20.59*)				**2.60%** (*15.97*)	**1.02%** (*10.05*)	**1.04%** (10.18)
LAR	**7.07%** (*25.70*)						
HL	**4.55%** (*20.88*)						
GGL	**1.56%** (*12.43*)						
No-Cue	**25.52%** (*43.71*)						

Figure 24 below shows the mean response time (left) and percent error (right) for each perfectly reliable single cue condition compared to the no-cue condition. A one-way repeated-measures ANOVA was used to examine the overall effect of cue type for both response time and percent error. The results revealed a significant main effect of cue type on response time, $F(3.16, 97.98) = 23.78$, $p < .001$, $\eta_p^2 = 0.43$. The second ANOVA revealed a significant main effect of cue type on percent error, $F(3.34, 103.41) = 12.33$, $p < .001$, $\eta_p^2 = 0.28$. Overall, the single, perfect cue conditions led to faster response times and lower error rates than the no-cue condition. It should be noted that these response times are, overall, substantially longer than in Experiment 1 (see Figure 5).

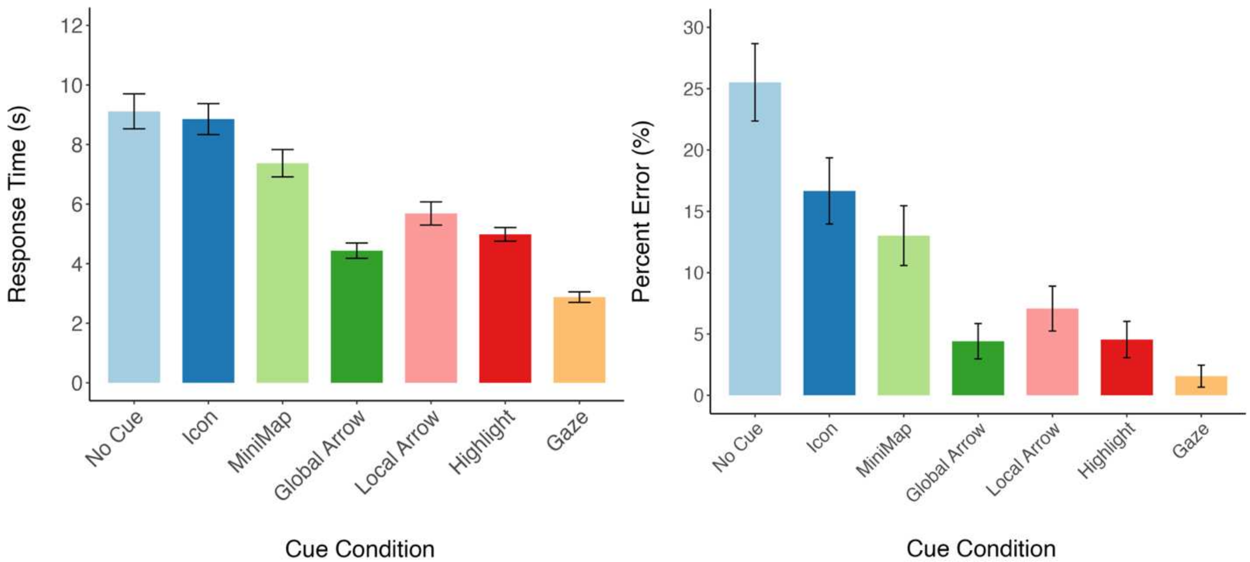

Figure 24. Experiment 2: The mean response time (left) and percent error (right) for all perfectly reliable single cues compared to the no-cue. Error bars represent one standard error of the mean.

Overall Benefits of a Cued Search Compared to an Unaided Search

Based on the results from Experiment 1, all the following analyses will exclude all the icon cues. The icon cue provides identity information about the target, making it a fundamentally different type of cue than the location cues. The location cues require the searcher to compare the cued target with a remembered image of the true target that degrades in working memory due to the passage of time between when they saw true target and when they located the target. The icon cue, on the other hand, continuously displays the image of the target and allows the searcher to complete a careful comparison of the image features of the cue with those of the real world object. Such a strategy will increase search time and improve accuracy. Therefore, due to the fundamentally different strategy employed for the icon cues, they will not be considered in the remaining analyses.

A one-way repeated-measures ANOVA was used to examine the overall effect of cueing for response time and percent error. Like Experiment 1, the cue conditions were coded as *cue* or *no-cue*. The factor *cue* was collapsed across all possible cue types and cue reliability (i.e., perfect, imperfect).

Response Time. The ANOVA results revealed a main effect of cueing, $F(1, 32) = 83.92$, $p < .001$, $\eta_p^2 = 0.73$, indicating faster response times when searching with cues ($M = 4.53$ s) than with no-cue ($M = 8.63$ s). These findings suggest the absence of cues significantly degrades the efficiency of search time.

Percent Error. Error data was analyzed the same way as response time except that percent error was included as the dependent measure. The results show that searching with cues (M = 7.45%) decrease error rate significantly more than searching with no-cue (M = 25.13%), F(1, 32) = 42.47, $p <$ .001, η_p^2 = 0.57. Again, the considerably worse performance than in Experiment 1 (25.13% vs 1348%) signals the greater difficulty of the search task, now with an even larger and more realistic search scene, albeit still static and two-dimensional.

Confirming Hypothesis 1, search tasks aided by an attention cue enhance search performance, allowing for an improvement of 4.1 s faster and 17.68% more accurate searches compared to an unaided search. These findings translate into a percentage increase of 90.29% in response time and 237.45% in percent error.

Effect of Cue Type and Reliability

The effect of reliability was examined the exact same way as Experiment 1: the no-cue and icon cue conditions were excluded from the analysis.

Response Time. The response time for imperfectly reliable cues (M = 4.83 s) did not significantly differ from the perfectly reliable cues, (M = 4.23 s; t(32) = 1.79, p = .08, 95% CIs [-0.08, 1.32], d = 0.32).

Percent Error. The percent error for imperfectly reliable cues (M = 11.20%) significantly increased compared to perfectly reliable cues, (M = 3.66%; t(32) = 8.33, $p <$ .001, 95% CIs [0.06, 0.09], showing an overall large benefit of higher reliability (Cohen's d = 1.40) in the cueing system.

Hypothesis 1A was supported for perfect error but not for response time. Unlike Experiment 1, there was no effect of reliability on response time. However, there was an effect of reliability on percent error.

Effect of Dual-Cueing

The following analysis examined the joint effects of cue type and reliability for single versus dual cues using the same analyses as Experiment 1. As a reminder, the greatest benefits of dual cueing are expected to occur with spatial location cues. Therefore, the only dual cues considered were those that combined a local cue (highlight, local arrow, gaze guidance line) with a global cue (minimap, global arrow). The effect of cue type (single versus dual cues) as a function of reliability was examined using a 2 (cue type) x 2 (cue reliability) repeated-measures ANOVA for both response time and percent error. The no-cue condition and all icon cue conditions were removed prior to the analysis.

Response Time. The mean response time data are shown in Figure 25. Dual cueing (M = 3.96 s) resulted in significantly faster response times than single cueing (M = 5.23 s), $F(1, 32) = 127.55$, $p < .001$, $\eta_p^2 = 0.80$. Neither the main effect of reliability, $F(1, 32) = 2.95$, $p = .095$, $\eta_p^2 = 0.08$, nor the interaction were statistically significant, $F(1, 32) = 3.42$, $p = .074$, $\eta_p^2 = 0.10$.

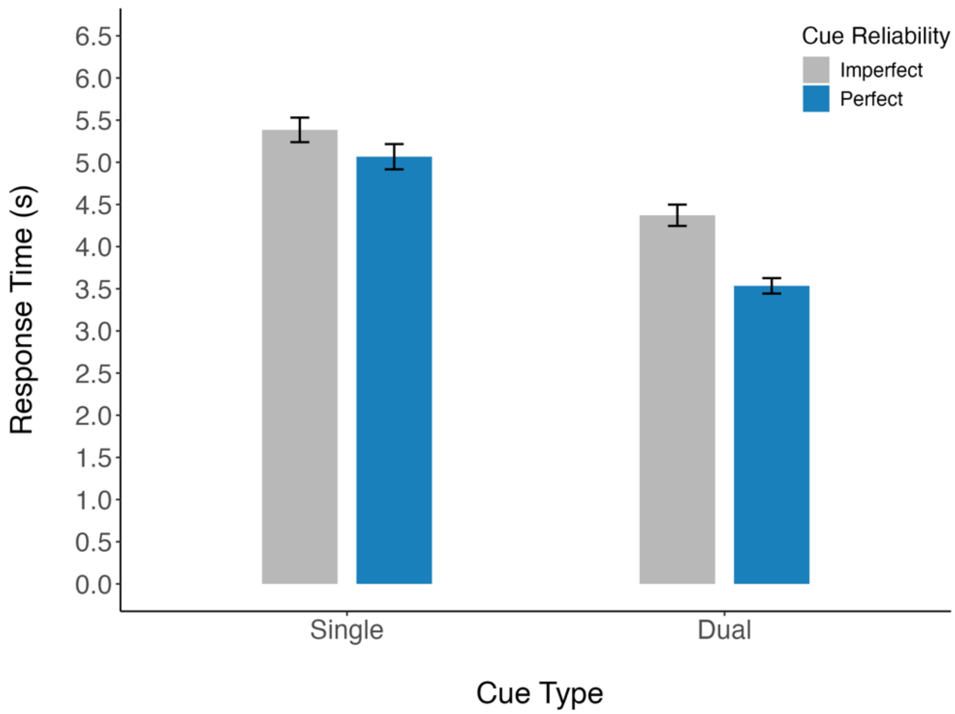

Figure 25. Experiment 2: The mean response time for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

Percent Error. Percent error was analyzed the same way as response time except that percent error was the dependent measure. The mean percent error data are shown in Figure 26. The results revealed that dual cueing (M = 6.35%) significantly reduced errors compared to single cueing (M = 8.76%), $F(1, 32) = 12.33$, $p = .001$, $\eta_p^2 = 0.28$. Additionally, imperfect cues (M = 11.20%) led to significantly greater errors than perfect cues (M = 3.66%), $F(1, 32) = 65.21$, $p < .001$, $\eta_p^2 = 0.67$. Critically, the interaction between cue type and reliability was significant, $F(1, 32) = 9.54$, $p = .004$, $\eta_p^2 = 0.23$. When cues were perfect, there were fewer errors with the dual cues (M = 1.62%) than the single cues (M = 6.10%; $t(32) = 4.38$, $p = .0001$, 95% CIs [0.02, 0.07], $d = 0.98$. But when cues were imperfect there were no significant differences in percent error ($t(32) = 0.44$, $p = .66$, 95% CI [-0.01, 0.02], d = 0.056). While this interaction was not significant in Experiment 1, the pattern of results are in the same direction: dual cueing only helps when reliability is perfect, whereas error is similar between dual and single cues when reliability is imperfect. This suggests that there is no additional cost of dual-cueing compared to single cueing when the automation is imperfect.

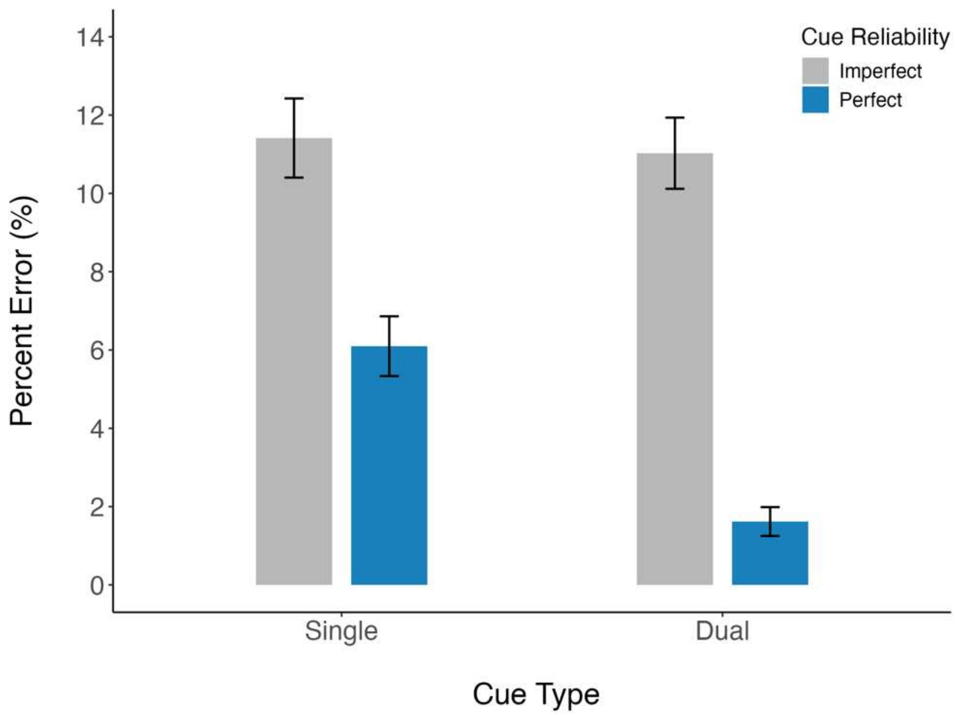

Figure 26. Experiment 2: The mean response time for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

In summary, dual cueing significantly improved visual search performance by reducing search time and error rates compared to single cues. The advantage of dual cueing was particularly enhanced when cues were perfect.

The Automation Bias: Correct vs Incorrect Automation Cueing

The automation bias was examined in the same way as in Experiment 1. A 2 (cue type: single versus dual) x 2 (automation performance: correct versus incorrect) repeated-measures ANOVA was conducted to examine the automation bias for response time and percent error.

Response Time. The mean response time for the automation bias is shown in Figure 27 below. The ANOVA results revealed a significant main effect of dual cueing, $F(1, 32) = 10.41$, $p = .003$, $\eta_p^2 = 0.25$, showing faster response times for dual cues ($M = 4.37$s) than single cues ($M = 5.38$ s). Additionally, response times were significantly faster when automation was correct ($M = 4.39$ s) than incorrect ($M = 7.03$ s), $F(1, 32) = 46.75$, $p < .001$, $\eta_p^2 = 0.59$. The interaction between cue type and automation performance was not statistically significant, $F(1, 32) < 0.001$, $p = .993$, $\eta_p^2 < .001$, precisely replicating the pattern of results in Experiment 1.

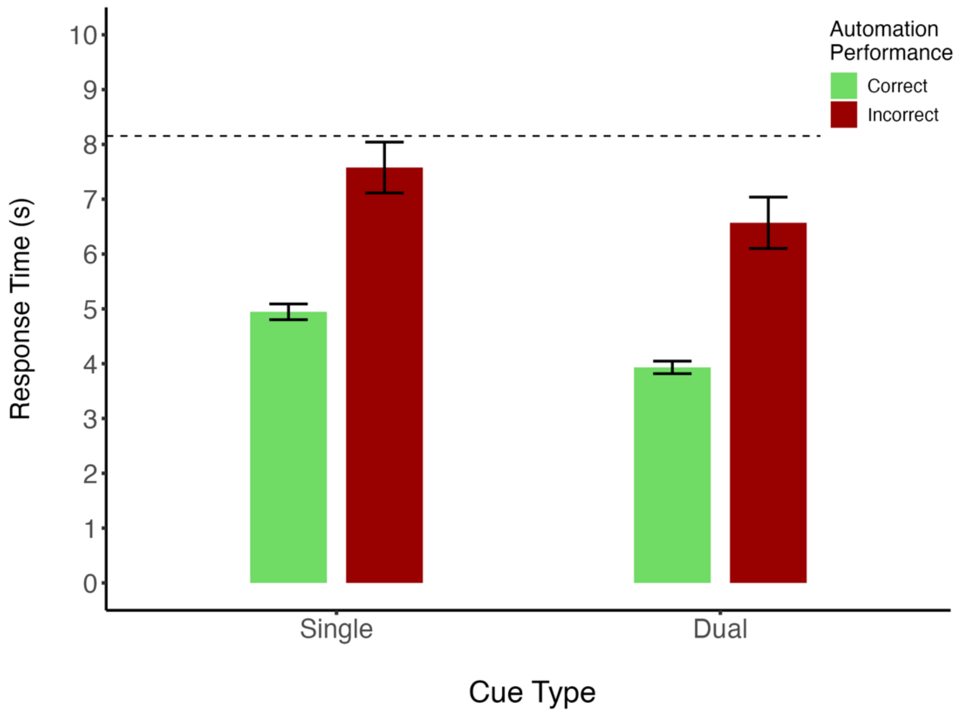

Figure 27. Experiment 2: The mean response time for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

Percent Error. The same analyses were conducted for percent error. The mean percent error data are shown in Figure 28. The main effect of dual cueing was not statistically significant, $F(1, 32) = 0.53$, $p = .47$, $\eta_p^2 = .02$. The main effect of automation performance was significant, $F(1, 32) = 65.92$, $p < .001$, $\eta_p^2 = 0.67$, revealing that percent error was greater when automation was incorrect ($M = 50.96$ %) than when correct ($M = 3.25$%), and much greater than the unaided condition ($M = 24.75$%). The interaction between cue type and automation performance was not significant, $F(1, 32) = 1.52$, $p = .23$, $\eta_p^2 = 0.05$, in contrast to the pattern observed in Experiment 1.

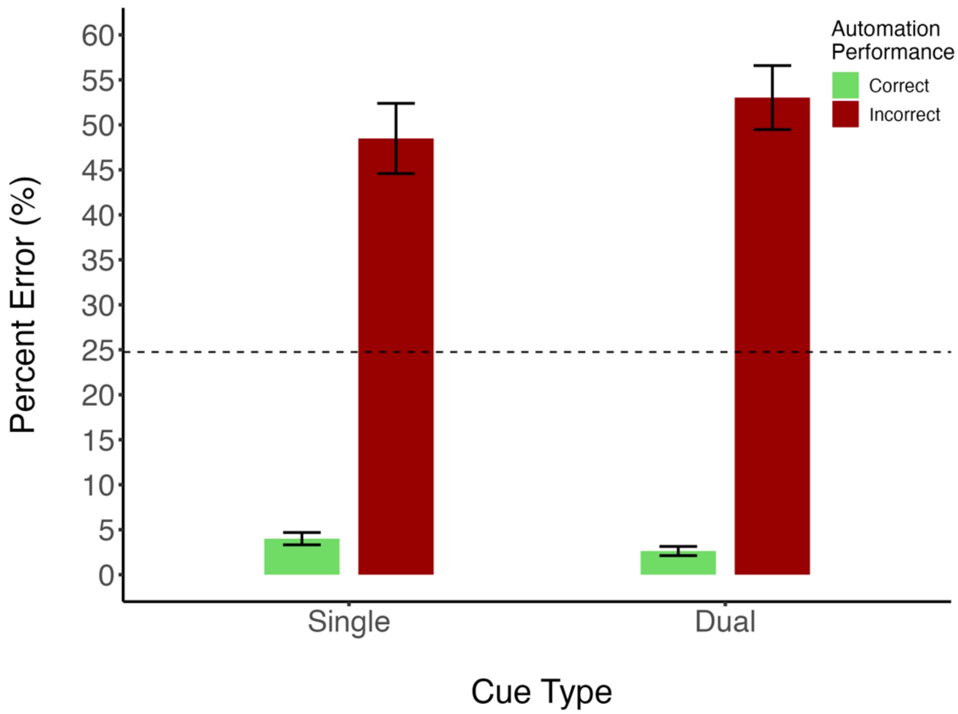

Figure 28. Experiment 2: The mean percent error for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

Prior literature found that the cues that resulted in the best performance when automation was correct, hurt performance the most when automation was incorrect (Yeh et al., 1999; Warden et al., 2023). Therefore, the following analysis examined the automation bias for the best and worst cues, specifically the gaze guidance line, global arrow, and minimap cues.

The mean response times are shown in Figure 29. Results from a 3 (cue type) x 2 (automation performance) repeated-measures ANOVA revealed a significant effect of cue type, $F(2, 64) = 4.05$, $p = .02$, $\eta_p^2 = 0.11$, showing significantly faster response times for the gaze cue ($M = 3.80$ s) than the global arrow ($M = 5.30$ s) or minimap ($M = 6.56$ s). Confirming the overall automation bias, response times were significantly faster when automation was correct ($M = 4.68$ s) than incorrect ($M = 7.93$ s), $F(1, 32) = 38.21$, $p < .001$, $\eta_p^2 = 0.54$. The interaction between cue type and automation performance was not significant, $F(1.87, 59.73) = 2.34$, $p = .105$, $\eta_p^2 = 0.07$.

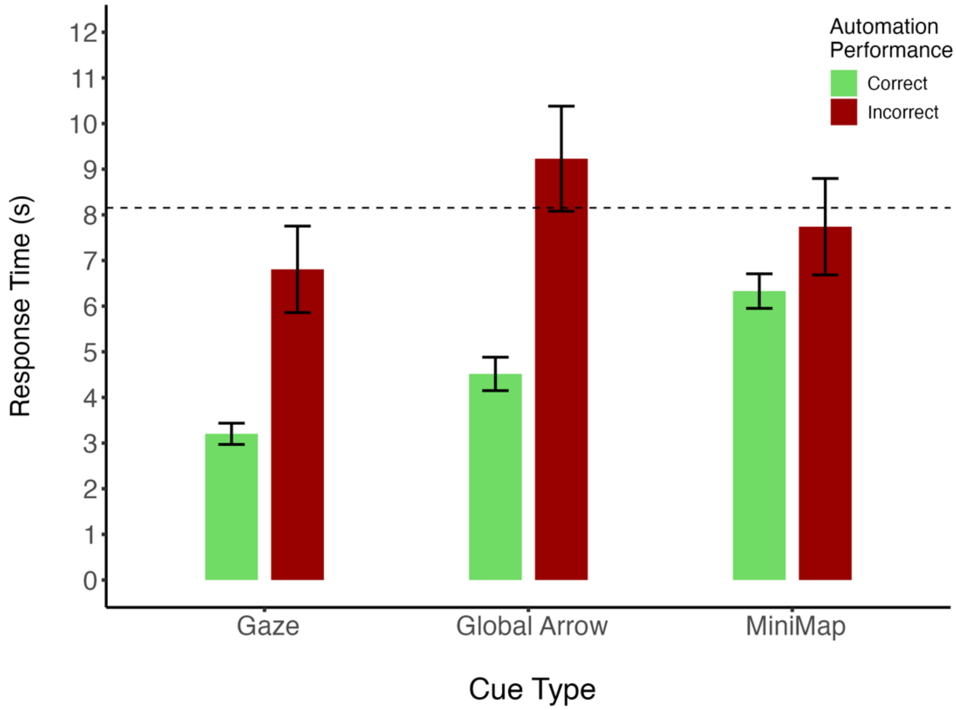

Figure 29. Experiment 2: The mean response time for single cue types (gaze, global arrow, and minimap) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

The following analysis again examines differences in automation bias for the gaze guidance line, global arrow, and minimap cues (see Figure 30) for percent error and was conducted the same way as for response time above. The ANOVA revealed a significant effect of automation performance, $F(1, 32) = 43.21$, $p < .001$, $\eta_p^2 = 0.57$, with percent error greater when automation was incorrect ($M = 47.47\%$) than correct ($M = 4.44\%$). Neither the effect of cue type ($F(1.91, 61.17) = 0.17$, $p = .845$, $\eta_p^2 < .001$) nor the interaction ($F(1.76, 56.38) = 1.936$, $p = .153$, $\eta_p^2 = 0.06$) were significant.

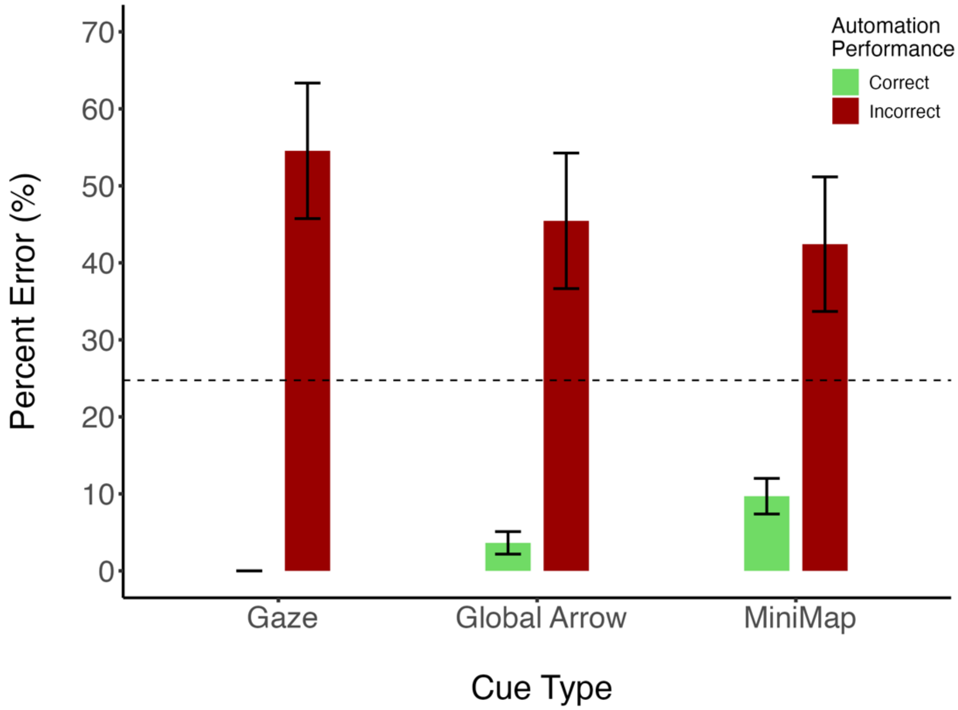

Figure 30. Experiment 2: The mean percent error for single cue type and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no-cue condition. Error bars represent one standard error of the mean.

Replicating the results from Experiment 1, participants exhibited an overall automation bias where they blindly followed the automation aid without checking the real-world data: cues led to higher error rates when the automation was incorrect compared to the unaided condition, and cued response times when the automation was wrong were never significantly less than the uncued condition. The extent of this automation bias neither differed for dual versus single cues nor for the best and worst single cues.

Effect of Cue Precision and Number of Cues: Redundancy Gain

The same exact analysis as Experiment 1 was used to examine the presence of a redundancy gain of the global-local cue combinations compared to either the global or local cues alone. Data were collapsed across cue reliability, and the no cue condition was excluded from the analysis.

Response Time. The mean response time across cue types is presented below in Figure 31. A one-way repeated measure ANOVA revealed a significant effect of cue type on response time, $F(1.32, 42.14) = 40.15$, $p < .001$, $\eta_p^2 = 0.56$. Pairwise comparisons (paired *t*-test) showed that response time was significantly slower for the global cues ($M = 5.90$ s) than the local ($M = 4.78$ s; $t(32) = -4.14$, $p = .0002$, 95% CIs [-1.69, -0.57], *d* = 0.60) and redundant cues (*M* = 3.96 s; *t*(32) = 8.46, *p* < .001, 95% CIs [1.48, 2.42], *d* = 1.02). Critically, in contrast to the findings of Experiment 1, redundant cues decreased response time more than local cues, *t*(32) = 6.56, *p* < .001, 95% CIs [0.57, 1.08], *d* = 0.47.

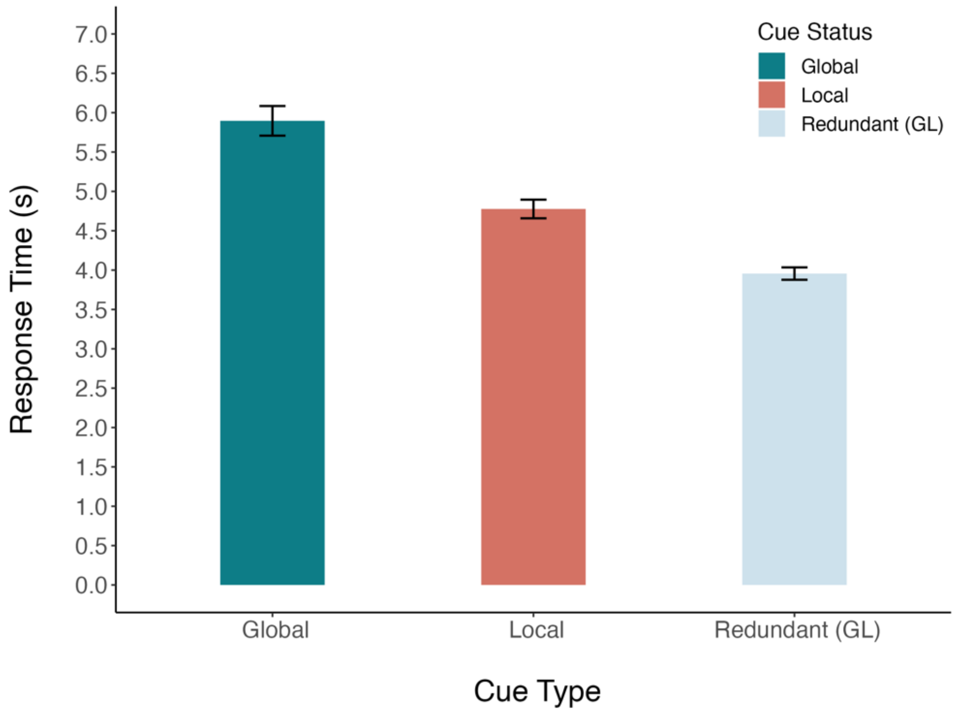

Figure 31. Experiment 2: The mean response time for cue type. Error bars represent one standard error of the mean.

Percent Error. The mean percent error across cue types is presented below in Figure 32. Results revealed a significant effect of cue type on percent error, *F*(1.36, 43.38) = 6.29, *p* = .003, η_p^2 = 0.16. Pairwise comparisons (paired *t*-test) showed the same pattern of results as the response time data. Specifically, percent error was significantly higher for global cues (*M* = 10.73%) than the local cues, (*M* = 7.45%; t(32) = -2.10, *p* = .044, 95% CIs [-0.06, -0.001], *d* = 0.39) and redundant cues (M = 6.35%; *t*(32) = 3.17, *p* = .003, 95% CIs [0.015, 0.071, *d* = 0.44). There was no significant difference between local and redundant cues, *t*(32) = -1.41, *p* = .167, 95% CIs [-0.005, 0.03], *d* = 0.26.

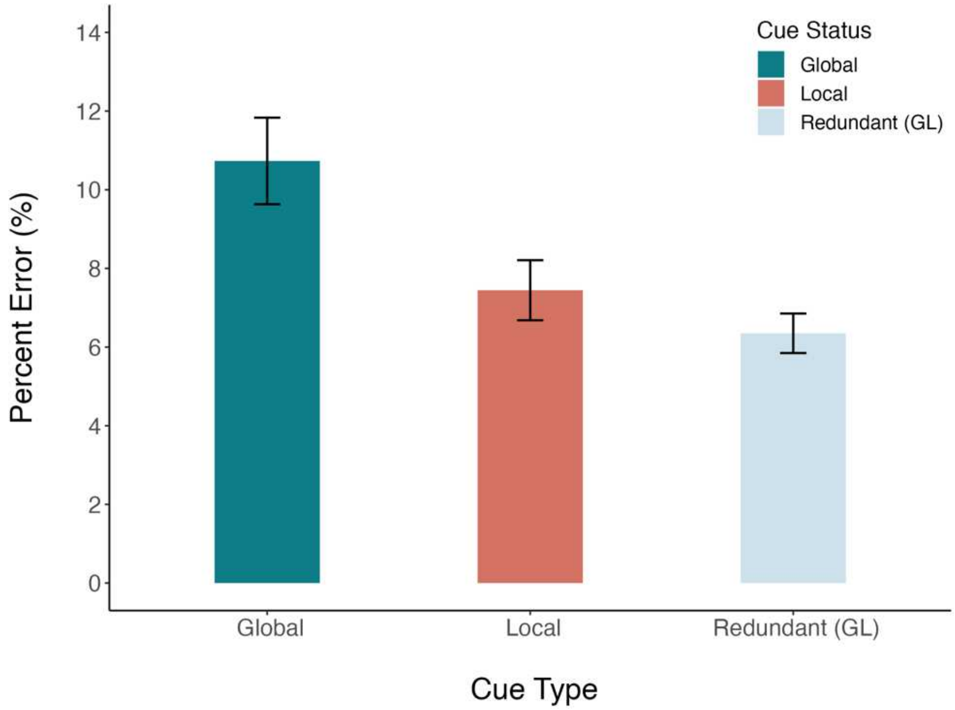

Figure 32. Experiment 2: The mean percent error for cue type. Error bars represent one standard error of the mean.

The pattern of results shows that redundant cues significantly reduced search time more than the global or local cues alone. Combining global and local spatial information in the redundant cue provides an advantage over either cue alone, thereby improving search time efficiency. However, for percent error, the results show that both local and redundant (global + local) cues improve overall search accuracy, suggesting no additional benefit of global cues.

Effect of Attentional Tunneling

The following analysis examines the effect of attentional tunneling for both response time and percent error. Only trials where a less expected high priority target appeared were included. For response time data, only trials where participants actually selected a high priority target were included because these are the only trials for which there is response time data for selecting that object. All data are collapsed across cue reliability and cue type. Instead, the analysis examined whether the cue status, either cued or uncued, of the routine target impacted the searcher's ability to find the high priority target.

Response Time. The mean response times are shown in Figure 33.

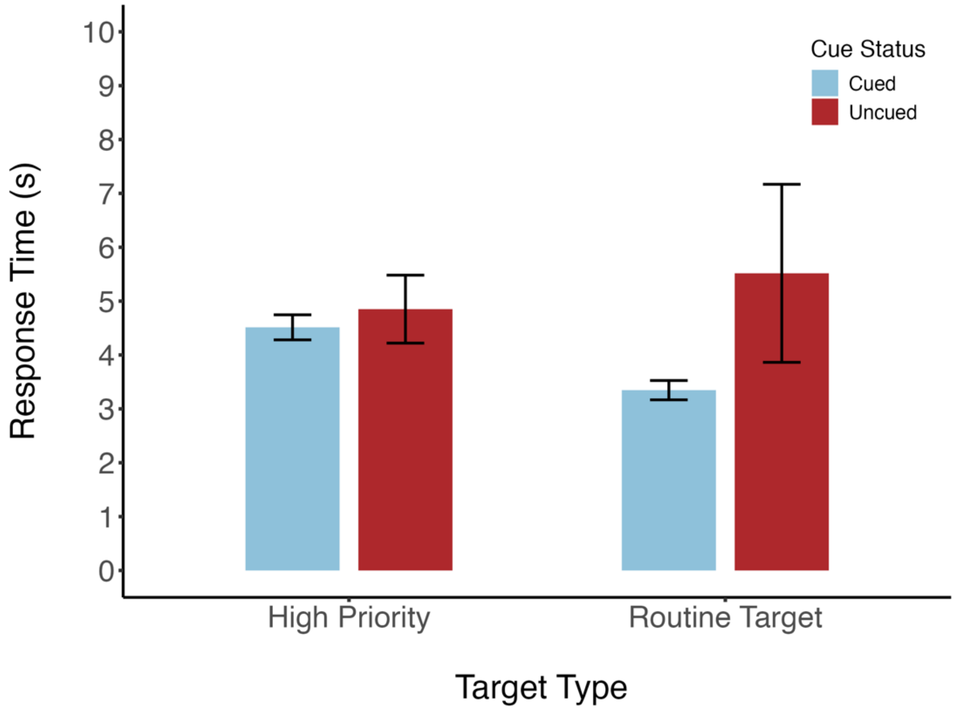

Figure 33. Experiment 2: The mean response time for target type (high priority and routine) as a function of whether the routine target was cued or uncued. Error bars represent one standard error of the mean.

A total of 8 participants were removed from the analysis because they failed to find the high priority target on specific trials resulting in missing data. A 2 (target type) x 2 (cue status) repeated measures ANOVA revealed a non-significant effect of target type on response time, $F(1, 17) = 0.33$, $p = .575$, $\eta_p^2 = 0.02$. Additionally, neither the main effect of cue status ($F(1, 17) = 2.28$, $p = .150$, $\eta_p^2 = .012$) nor the interaction ($F(1, 17) = 0.65$, $p = .431$, $\eta_p^2 = 0.04$) were significant. It should be noted that the high amount of variability in the uncued condition for the high priority target is due to the limited number of trials for which the less expected, high priority target could appear. The other source of variability for both uncued targets may be due to the different search strategies employed by participants when a target was uncued.

Percent Error. Percent error data are shown in Figure 34. The ANOVA revealed a significant main effect of target type (routine, high priority), $F(1, 32) = 120.80$, $p < .001$, $\eta_p^2 = 0.79$, indicating that percent error is greater for the less expected high priority target (M = 64.50%) compared to the routine target (M =

10.54%). Neither the effect of cue status ($F(1, 32) = 0.43$, $p = .518$, $\eta_p^2 = 0.01$) nor the interaction ($F(1, 32) = 1.17$, $p = .288$, $\eta_p^2 = 0.03$) were significant.

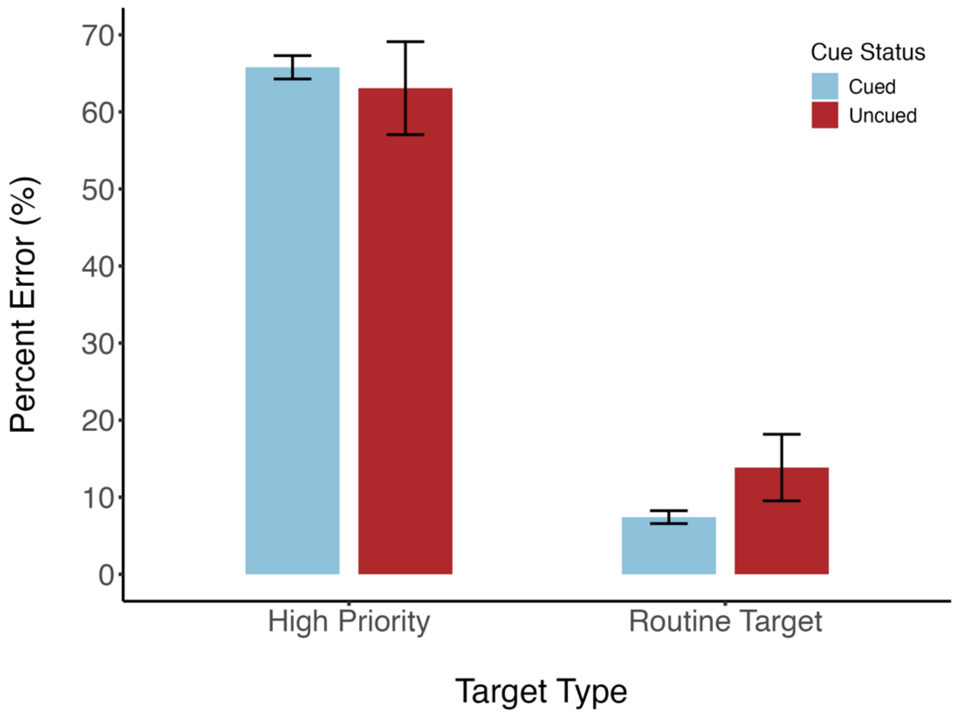

Figure 34. Experiment 2: The mean percent error for target type (high priority and routine) as a function of whether the routine target was cued or uncued. Error bars represent one standard error of the mean.

To further examine the effects of attention tunneling and whether dual cueing induces greater attention tunneling than single cueing, the data were further categorized based on target type (high priority, routine) and cue status (no cue, single cue, dual cue). Figure 35 shows mean response time for high priority and routine targets as a function of dual versus single cueing.

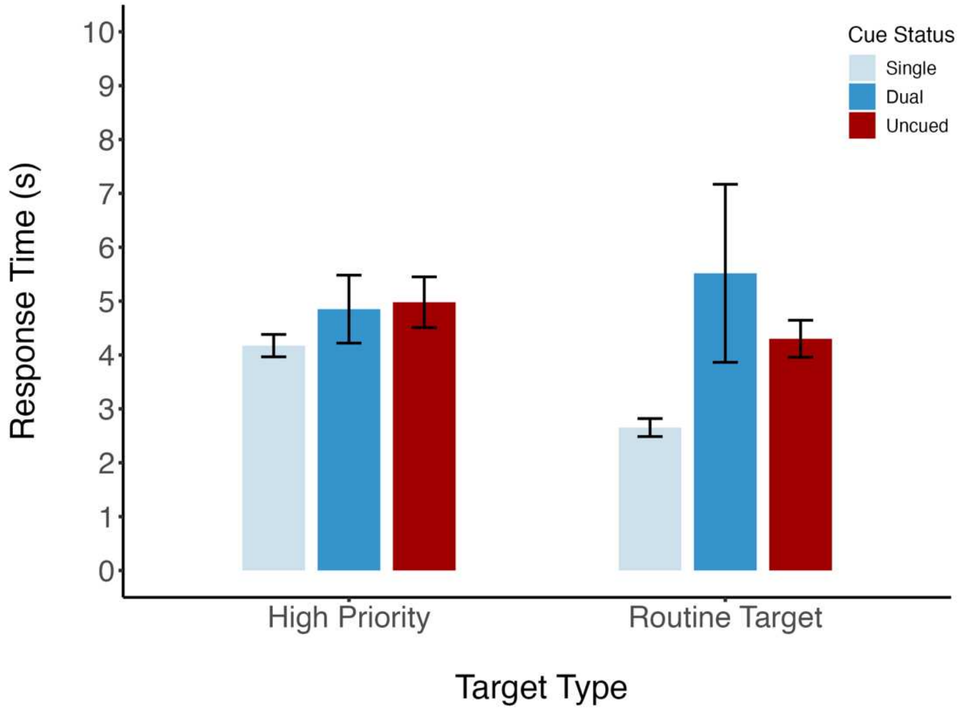

Figure 35. Experiment 2: The mean response time for target type (high priority and routine) as a function of cue status (dual, single, uncued). Error bars represent one standard error of the mean.

The ANOVA repealed a significant effect of target type (F(1, 17) = 11.90, p = .003, η_p^2 = 0.41), showing that response time was greater for the routine target (M = 7.47 s) than the less expected high priority target (M = 5.21). There was also a significant effect of cue status (F(1.51, 25.60) = 3.74, p = 0.049, η_p^2 = 0.18), indicating faster response times with the dual cue (M = 4.86 s) than the single (M = 7.00 s) and uncued (M = 7.15 s) conditions. The interaction was not significant (F(1.27, 21.55) = 2.512, p = .122, η_p^2 = 0.13).

Figure 36 shows mean percent errors for high priority and routine targets as a function of dual versus single cueing. The ANOVA revealed a significant main effect of target type, F(1, 32) = 143.51, p < .001, η_p^2 = 0.82, indicating that percent error was greater overall for the less expected high priority target (M = 64.58%) compared to the routine target (M = 9.91%). Neither the main effect of cue status (F(1.34, 42.92) = 0.43, p = .575, η_p^2 = 0.01) nor the interaction (F(1.37, 43.87) = 1.70, p = .17, η_p^2 = 0.05) were significant.

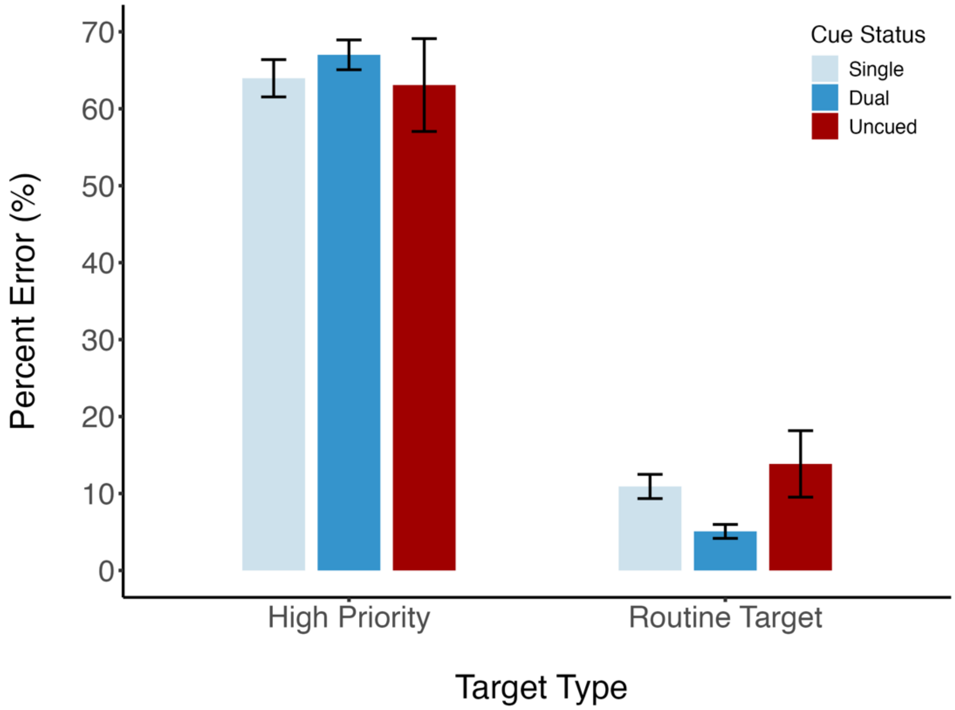

Figure 36. Experiment 2: The mean percent error for target type (high priority and routine) as a function of cue status (dual, single, uncued). Error bars represent one standard error of the mean.

Similar to Experiment 1, there was weak overall evidence for attentional tunneling. Overall, errors for detecting the high priority target show a pattern of being greater when the routine is cued compared to uncued, but none of these differences were significant. Across these analyses, it seems like there is a speed-accuracy tradeoff where people are quick to find the less expected high priority target, but this comes at a cost to accuracy. Like Experiment 1, dual-cueing did not enhance the effects of attentional tunneling.

Effect of Cue Display Imagery: World vs Screen Referenced Coordinates

Cue effectiveness in augmented reality is contingent upon whether the cues are displayed in world-referenced or screen-referenced coordinates. As a reminder, world-referenced coordinates (i.e., conformal imagery) have a one-to-one correspondence with the exact (x, y) or (x, y, x) position of information in the real-world scene behind the display, meaning that the imagery is linked or overlays the real world information. Screen-referenced coordinates (i.e., nonconformal imagery) are linked to the (x, y) coordinates of the display itself, meaning that it always moves with the user as they move their head. To examine the effects of cue display imagery, spatial cues only were categorized as either world-referenced (i.e., global arrow, local arrow, highlight, gaze guidance), screen-referenced, (i.e., minimap), or a combination of the two in the cause of dual cues. All icon cues and the no-cue conditions were excluded from the analysis.

To first examine the overall effect of cue display imagery for single cues, a one-way repeated-measures ANOVA was used for response time and percent error. The mean response time data are shown in Figure 37.

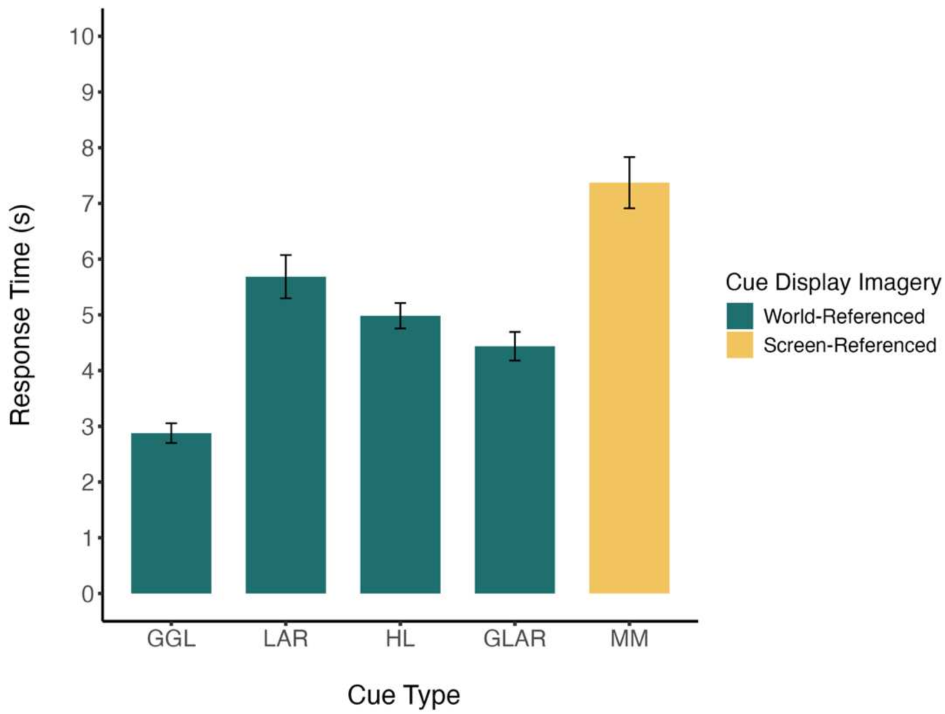

Figure 37. Experiment 2: Mean response time as a function of single cues that are either world-referenced (teal) or screen-referenced (yellow). Error bars represent one standard error of the mean.

Results revealed a significant main effect of cue imagery, $F(2.8, 86.83) = 19.412$, $p < .001$, $\eta_p^2 = 0.38$. Pairwise comparisons (paired *t*-test) revealed that, of the local, world-reference cues, the gaze guidance line ($M = 2.88$ s) led to the fastest response time compared to all other cues (*ps* < .001). The global, screen-referenced minimap ($M = 7.37$ s) led to the slowest response time compared to all other cues (*ps* < .02).

Percent error data was analyzed the same way as response time (see Figure 38).

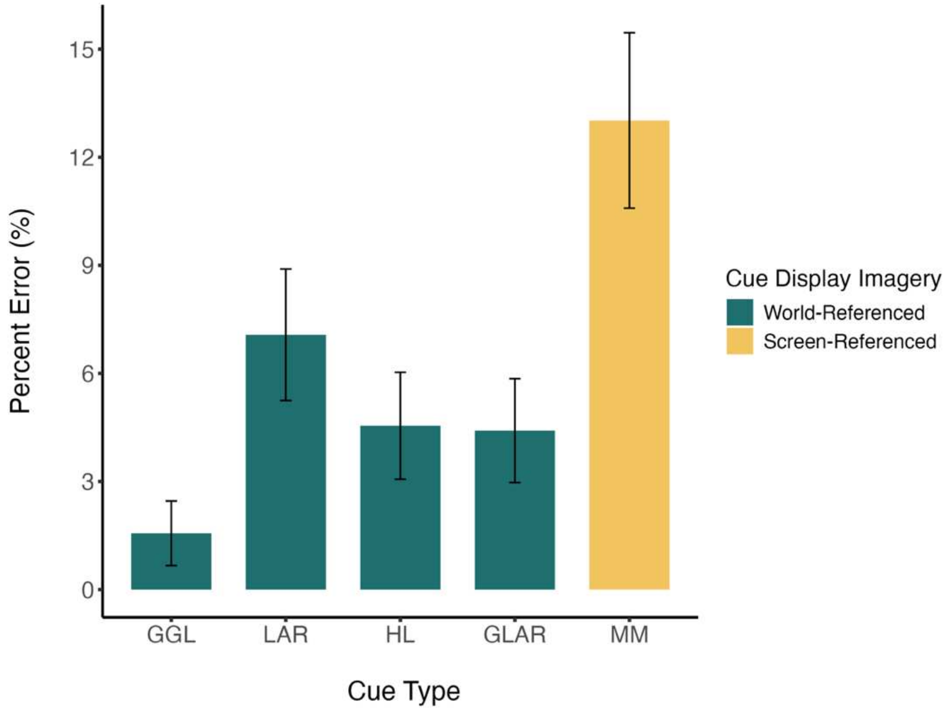

Figure 38. Experiment 2: Mean percent error as a function of single cues that are either world-referenced (teal) or screen-referenced (yellow). Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue type, $F(2.68, 82.98) = 3.53$, $p = .022$, $\eta_p^2 = 0.10$. Pairwise comparisons show that the local, world-reference gaze guidance line reduced error rate more than the local arrow and minimap cues (*ps* < .02). There were no differences between the gaze guidance line, highlight, and global arrow cues (*ps* > 0.18). The minimap led to the greatest percent error compared to the global arrow, highlight, and gaze guidance cues (*ps* < .054) but not the local arrow cue ($p = .18$).

Next, the effect of cue display imagery was examined for both single and dual cues. Cue display imagery was coded as either world-referenced (WR), screen-referenced (SR), world-world referenced (W-WR), or screen-world referenced (S-WR). The goal of this analysis is to assess the impacts of combining display imagery relative to single display imagery. The mean response times are shown in Figure 39.

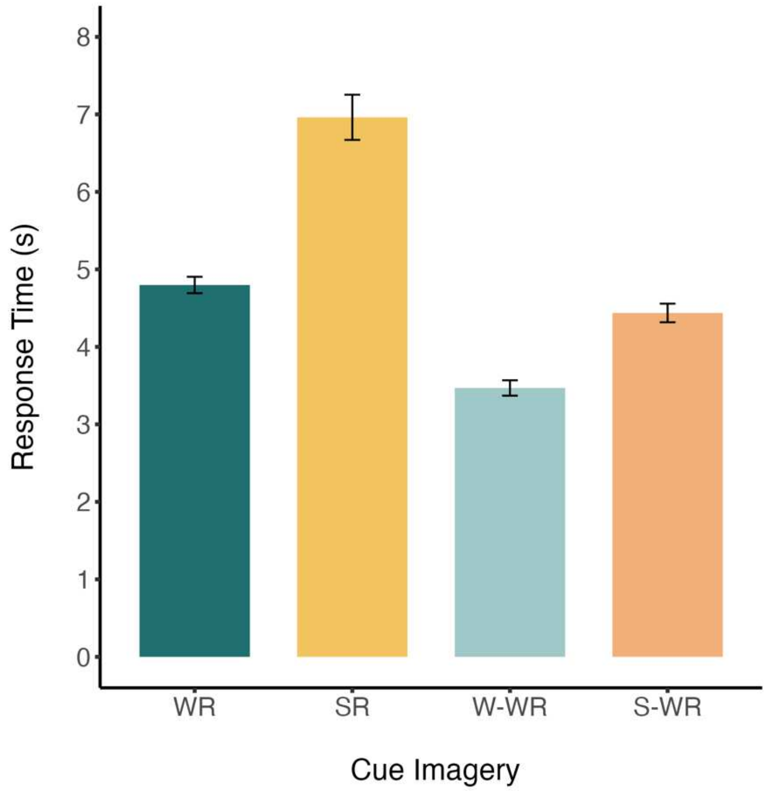

Figure 39. Experiment 2: Mean response time as a function of cue imagery for world-referenced (WR), screen-referenced (SR), world-world referenced (W-WR) and screen-world referenced (S-WR) cues. Error bars represent one standard error of the mean.

A one-way repeated-measures ANOVA revealed a significant main effect of cue display imagery, $F(1.20, 38.39) = 44.12$, $p < .001$, $\eta_p^2 = 0.58$. Pairwise comparisons show that world-world referenced cues led to significantly faster response times compared to all other cues (p*s* < .001), and, most critically, faster than the screen-world reference cue ($t(32) = 7.21$, $p < .001$, 95% CI [0.70, 1.25], $d = 0.56$). The screen-referenced cue resulted in slower response times compared to all other cues (*ps* < .001).

Percent error was analyzed the same way as response time (see Figure 40 below).

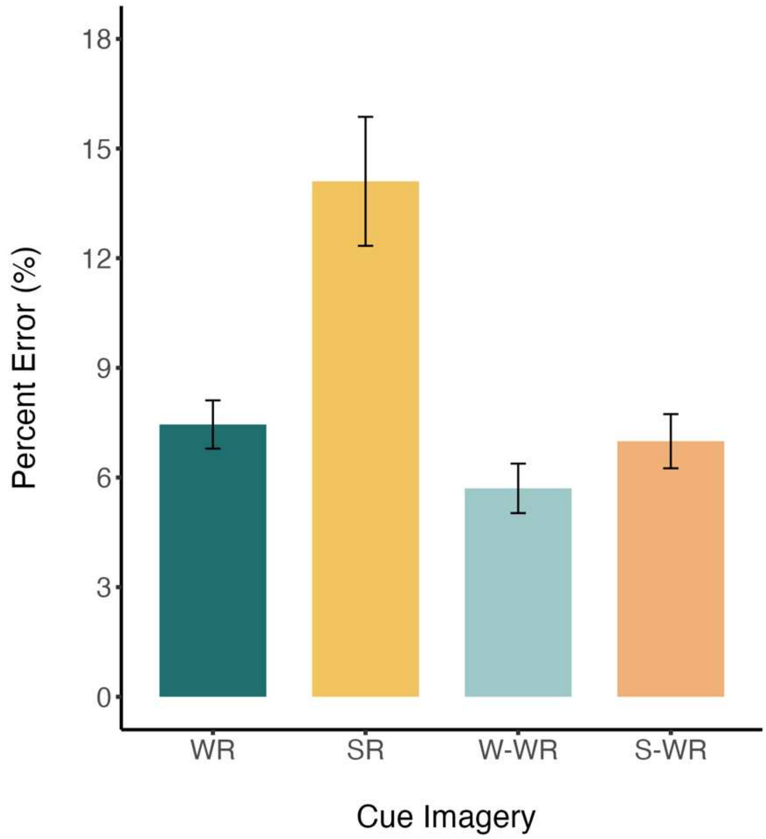

Figure 40. Experiment 2: Mean percent error as a function of cue imagery for world-referenced (WR), screen-referenced (SR), world-world referenced (W-WR) and screen-world referenced (S-WR) cues. Error bars represent one standard error of the mean.

The ANOVA revealed a significant main effect of cue imagery, $F(1.19, 38.05) = 8.99$, $p < .001$, $\eta_p^2 = 0.22$, and showed a similar profile to the response time Like response time, pairwise comparisons show that the W-WR cues reduced error rate more than the S-WR cues ($t(32) = 2.24$, $p = .03$, 95% CI [.001, .02], $d = 0.30$) and more than all of the other cues (*ps* < .012). Also, like response time, the screen-reference cue resulted in the highest error rate compared to all other cues (*ps* < .002).

In summary, the best cue is always the dual-cue that uses world-reference imagery for both cue types (i.e., global arrow combined with either the gaze guidance line, local arrow, or highlight). This finding reiterates the earlier finding that dual-cueing is best for attention guidance, especially when both of those cues are in world-referenced coordinates.

Field of View Analyses

A 2 (FOV) by 2 (cue precision) repeated-measures ANOVA was conducted to examine the influence of global, local, and combined global-local cues on performance when the cued target was either inside or outside of the field of view (FOV). Cue precision here refers to whether the cue provided global, local, or combined global-local spatial information. The FOV was categorized as whether the cued target was inside or outside the immediate HMD FOV at the start of the trial. If the cued target was outside the FOV it could not be seen until head rotation brought it into the FOV. The mean response time data are shown in Figure 41 below.

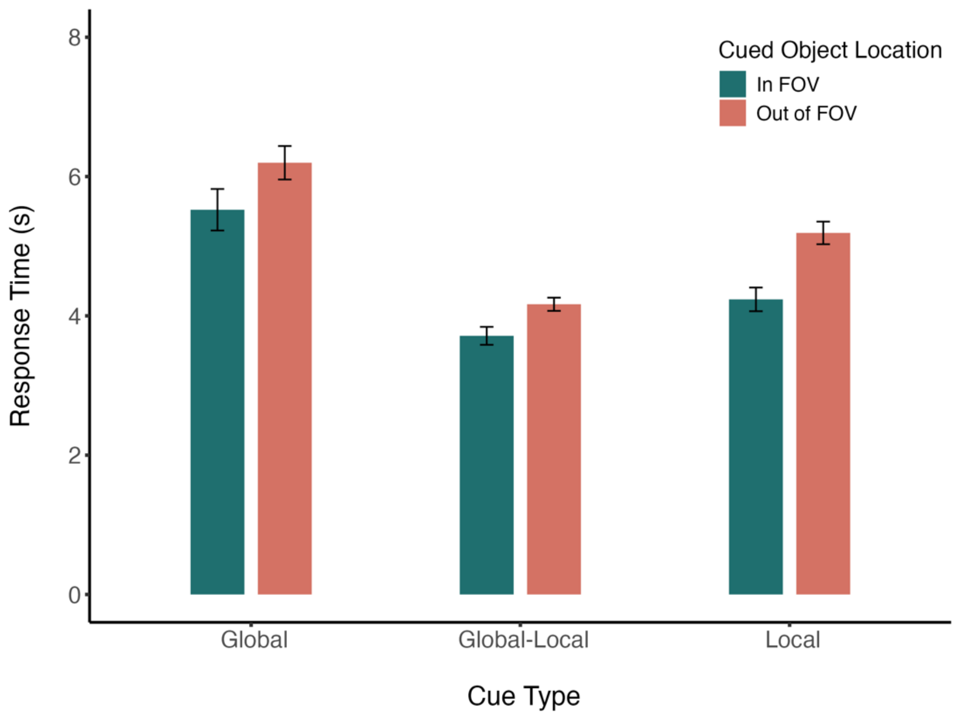

Figure 41. Experiment 2: Mean response time data for cue type (global, local, or global-local) as a function of whether the cues were in (green) or out (red) of the immediate field of view of the device. Error bars represent one standard error of the mean.

Response Time. The ANOVA showed a significant effect of cue precision seen before, $F(1.33, 42.62) = 37.62$, $p < .001$, $\eta_p^2 = 0.54$, showing that the combined global-local cues resulted in the fastest response times ($M = 3.96$ s) compared to the local ($M = 4.78$ s) and global ($M = 5.90$ s) cues. There was also a significant effect of FOV, $F(1, 32) = 10.05$, $p = .003$, $\eta_p^2 = 0.24$, indicating that response times were faster when the cued target was inside the FOV (*M* = 4.18 s) than outside (*M* = 4.83 s). The interaction was not significant, $F(1.79, 57.25) = 0.47$, $p = .60$, $\eta_p^2 = 0.01$.

Percent Error. Percent error was analyzed the same way as response time data (see Figure 42 for means).

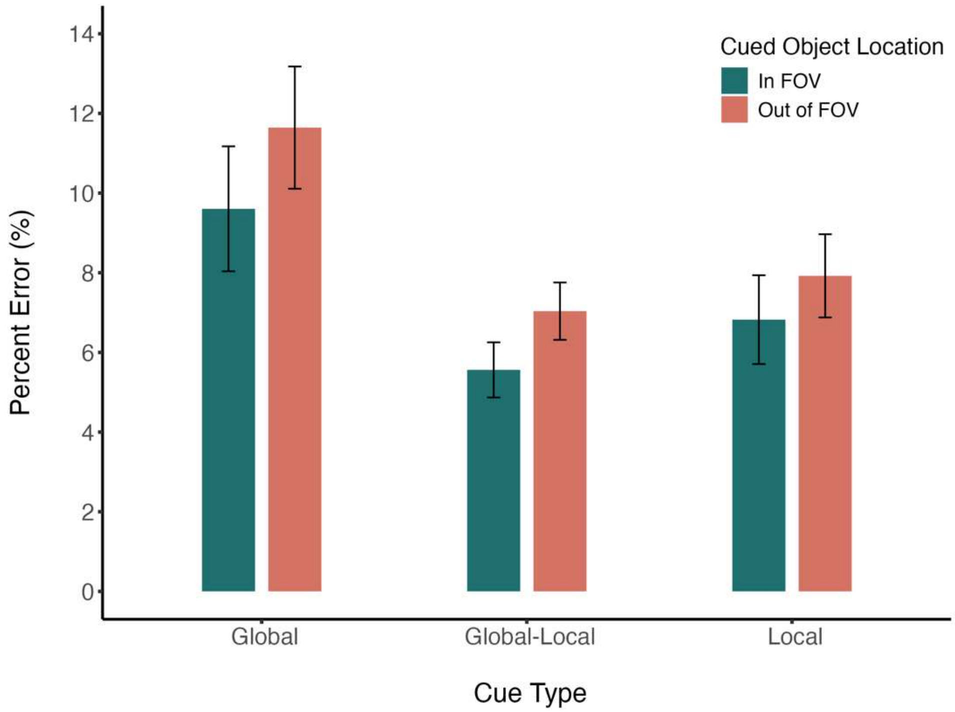

Figure 42. Experiment 2: Mean percent error data for cue type (global, local, or global-local) as a function of whether the cues were in (green) or out (red) of the immediate field of view of the device. Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue precision, $F(1.35, 43.07) = 5.67$, $p = .014$, $\eta_p^2 = 0.15$, indicating that percent error was reduced the most with global-local cues (M = 6.35%) compared to local (*M* = 7.45%) and global (*M* = 10.73%) cues. Neither the effect of cue FOV ($F(1, 32) = 1.89$, $p = .179$, $\eta_p^2 = 0.06$) nor the interaction ($F(1.41, 44.96) = 0.06$, $p = .891$, $\eta_p^2 = 0.002$) were significant.

Overall, for response time, the findings further support the benefits of dual cueing, showing that the combined global-local cue improves search time more than either cue alone, particularly true when the target is outside the immediate FOV.

Discussion

The current experiment examined the influence of attention cueing during a large static visual search task using attention cues presented with an AR-HMD. Replicating the findings of Experiment 1, H_1 was supported, showing an overall performance benefit for cueing compared to an unaided search. The cueing benefit was significantly greater for Experiment 2 (AR-HMD search task) than Experiment 1 (desktop search task) for response time ($t(78) = -5.03$, $p < .001$, $d = 1.08$), and percent error ($t(78) = 2.89$, $p = .002$, $d = 0.66$). The desktop search task benefits from the use of peripheral vision, a feature unavailable in the AR-HMD where target objects outside of the FOV are invisible until head movements are invoked to bring information into the FOV. Additionally, replicating the findings from Experiment 1, dual-cueing helped performance more than single cueing, supporting hypothesis H_2, and this dual cue benefit, to response time, was greater in Experiment 2 (Figures 24 and 25) than in Experiment 1 (Figures 8 and 9).

Unlike Experiment 1, imperfect cues did not significantly degrade response time performance compared to perfect cues. However, imperfect cues did hurt accuracy, indicative of a true automation bias and replicating findings from Experiment 1. Additionally, the advantage of dual cues was found to be greater for both response time and percent error when cues were perfect. Unlike Experiment 1, there was no main effect of reliability for response time. However, the pattern of results is in the same direction, suggesting faster search times when automation is perfect compared to imperfect. There was, however, a main effect of reliability for percent error, suggesting that participants may have blindly followed the automation's recommendation when the automation was incorrect regardless of whether the cue was single or dual. Results concerning the automation bias replicated those from Experiment 1, showing an overall tendency for searchers to blindly follow the advice of the automation when it was wrong. The extent to which people exhibited the automation bias was not moderated by the type of cue as had been shown in some prior research (Warden et al., 2023), such that the best cue when automation was perfect did not result in significantly worst performance when the automation erred. Cues, in general, caused an automation bias independent of their effectiveness. However, it should be noted that there was a non-significant trend suggesting that dual cues may lead to more automation bias than single cues. Further exploration is needed to assess whether dual cues enhance the automation bias.

Similar to Experiment 1, there was little evidence of attentional tunneling. In general, participants seemed to trade off accuracy for speed when searching for the less expected high priority target. Dual-cueing did not further enhance attentional tunneling, contradicting H_{2B}. However, there was an overall trend suggesting that cueing the routine target, in general, led to higher errors when detecting the high priority target compared to when the routine target was uncued.

Furthermore, the results also showed that redundant cues that combine global and local information reduced search time the most compared to either single cue alone, indicating an overall benefit of redundancy for search time (Figure 31). When searching with the AR-HMD for a target that falls outside of the immediate FOV of the device, having the dual cue that provides global and local information becomes helpful because it signals to the user which direction to turn their head to find the local cue. For accuracy, there was no such redundancy gain. These findings suggest that dual cueing provides a redundancy gain that benefits time-critical task, such as detecting a hazard quickly without any cost to accuracy.

The current results found support for H_{4A}, which predicted that the most compelling cues (i.e., those using world-referenced coordinates) result in better performance compared to screen-reference cues. The current results show that dual-cues where both cues use world-referenced coordinates led to superior response time and accuracy compared to all other cues. However, in contrast to prior findings (e.g., Warden et al., 2023; Yeh & Wickens, 2001) there was no evidence suggesting that the most compelling cue (i.e., gaze guidance line) induced a greater degree of automation bias and cognitive tunneling, contradicting H_{4B}.

EXPERIMENT 3: DYNAMIC VISUAL SEARCH TASK IN VIRTUAL REALITY

Description

The purpose of Experiment 3 is to examine cueing effectiveness during a dynamic visual search task presented within a realistic 3D virtual environment where participants simulate walking forward along a linear path. Another purpose of this experiment is to assess whether findings from Experiment 1 and 2 generalize to a dynamic and more realistic visual search task within a 3D environment. Specifically, whether cue effectiveness, dual-cueing, and frame-of-reference cueing results generalize to a dynamic wide field of view search task.

Method

Participants. A total of 44 participants were recruited from the psychology research pool at Colorado State University and participated in the experiment in exchange for course credit. No participants from Experiment 1 or Experiment 2 participated in Experiment 3. All participants had self-reported normal or corrected-to-normal vision and were screened for colorblindness using an electronic version of Ishihara's test. No demographic information was collected.

Stimuli and Apparatus. Participants completed the search task using the Oculus Meta Quest 3, which simulated a real-world 3D environment. The entire experiment was created in the game engine Unity (version 2020.3.28f1FOV) and used the mixed reality toolkit. The environment consisted of a linear dirt path that was 136.67 feet long (1640 inches). The real-world environment had a similar terrain to the previous experiments, except that it now simulating a 3D terrain (see Figure 43 below for an illustration).

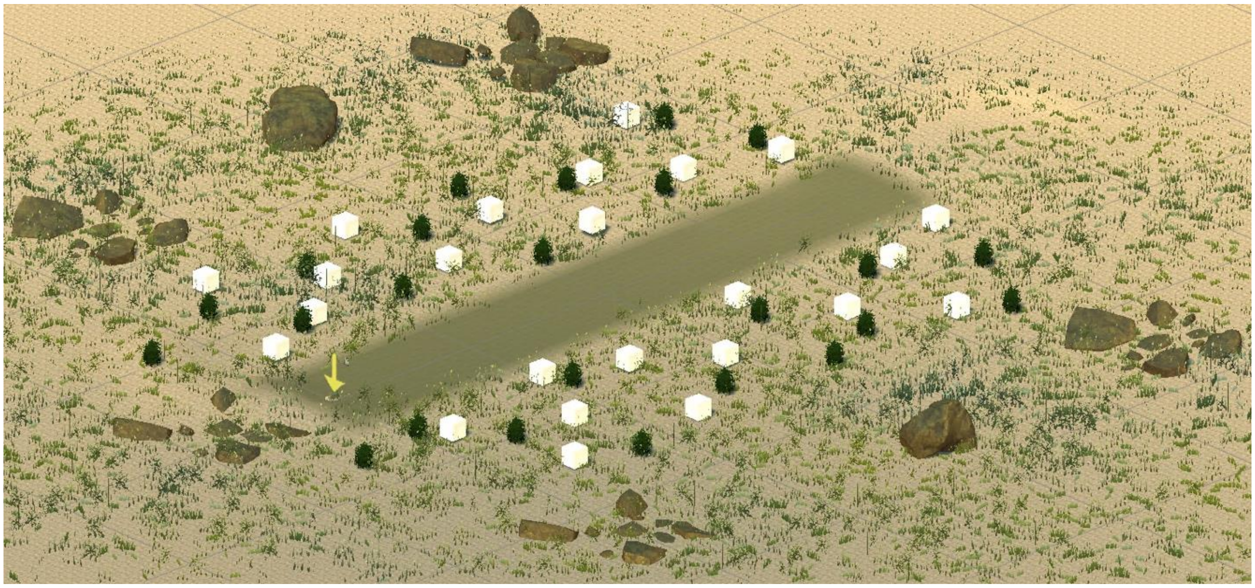

Figure 43. A side-view of the search scene created in Unity. This illustrates the type of foliage present in the scene. The yellow arrow is where the participant started on each trial.

The scene consisted of 12 possible object locations presented on either side of the path for a total of 24 possible object locations (see Figure 44).

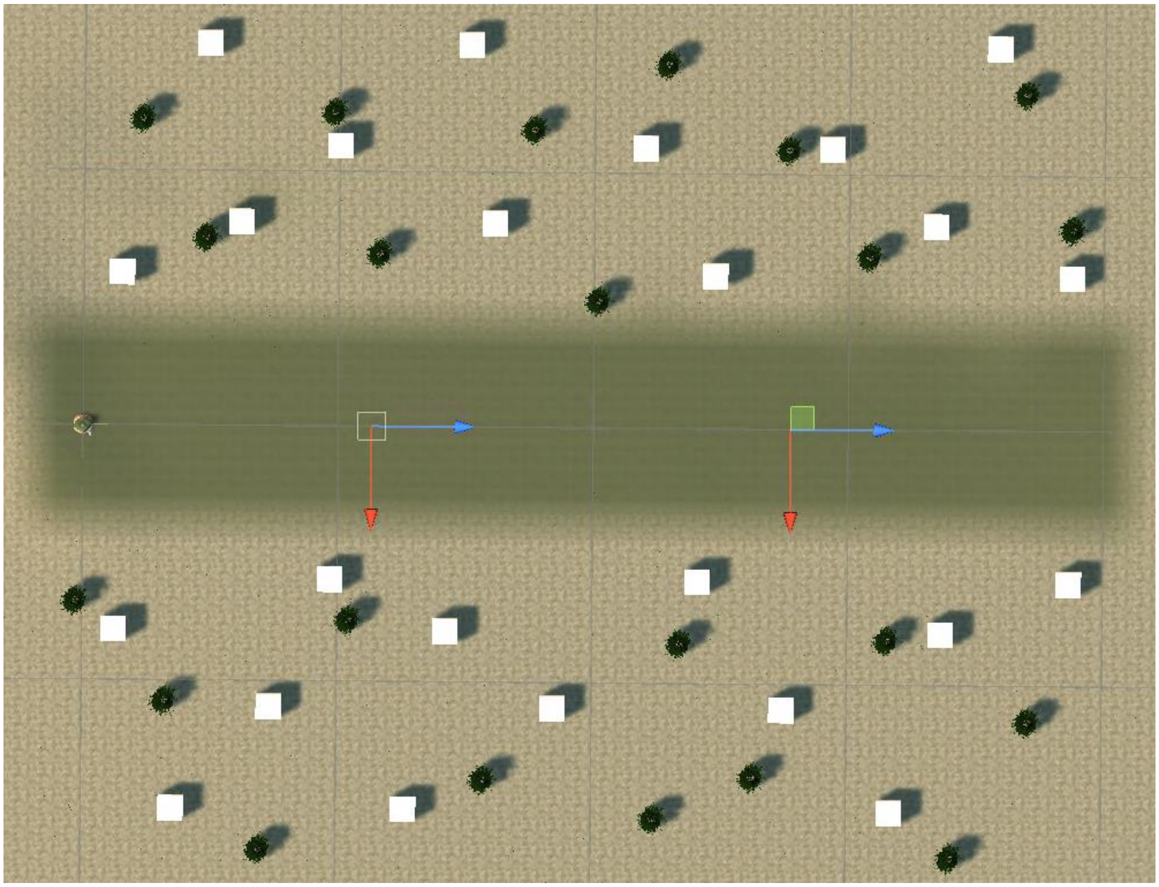

Figure 44. A top-down view of the search scene created within the Unity game engine. The white boxes represent the possible locations where objects could be positioned. The green path represents the linear path where forward movement starting from the left and moving to the right was simulated.

At the start of each trial, eleven objects were uniformly distributed on either side of the path, for a total of 22 objects in the search scene. The 22 objects were randomly pulled from a set of 40 routine objects. Figure 45 below shows an example of each type of routine object and the four high priority objects. All objects were 3D and were acquired from the Unity asset store. Because the search scene was so large, objects were located 152 inches apart from one another.

Figure 45. Examples of each type of routine object category (top) and all five of the less expected, high priority objects (bottom). These example images are 2D but the objects in the virtual search scene were 3D.

Participants moved forward along the path as they searched for objects. There was zero acceleration as they moved through the VR simulation to minimize motion sickness. The speed at which participants moved through the scene was constant and set to a normal walking speed of 5.58 feet per second (i.e., 1.7 unity measures per second) and the full path was traversed in an average of 24 seconds from start to finish without engaging in any searching. The height of the participant in the virtual environment 70.87 inches (1.8 unity measures) tall. Within the VR experiment, attention cueing aids were presented so that they simulate the user wearing an AR-HMD, meaning that they were either directly in the line of sight of the user or they were embedded in the scene with simulated augmented reality properties (see Figure 46).

Figure 46. A screen capture from Unity illustrating the forward field-of-view from the participants perspective in the virtual scene. The transparent box with the fixation cross is where the attention cues could be displayed.

Below, in Figure 47, are examples of the search scene from the perspective of the participant where the actual 3D objects are visible.

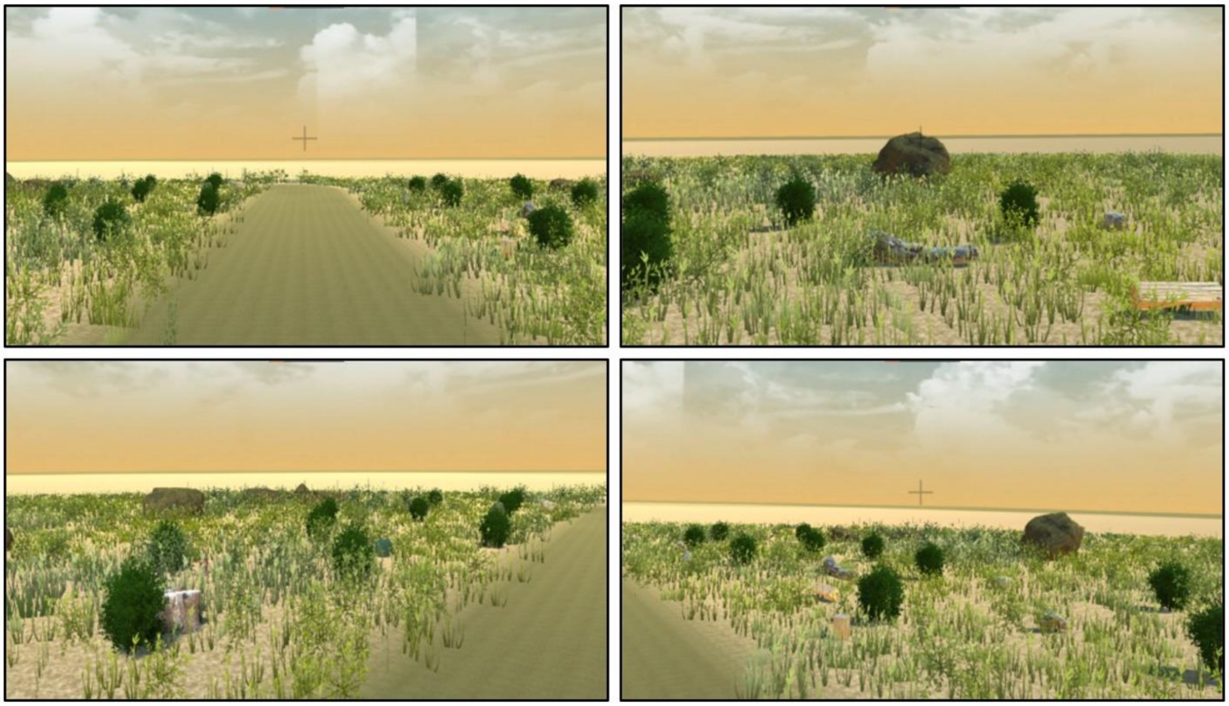

Figure 47. Four different views from the user's first person perspective in the 3D VR search environment.

The present experiment included 12 cue conditions, including the no cue condition. The specific single spatial location cues included global cues (global arrow, two minimap cues) and local cues (gaze guidance line, local arrow). The minimap cues differed in their location on the simulated AR-HMD, being either at the center of the display or 20 degrees downward from the center of the simulated HMD. The dual-cue conditions consisted of one global cue and one local cue. Like Experiment 2, the display imagery of the cueing aids was either world-referenced (i.e., global arrow), screen-referenced (i.e., minimap), or, in the case of dual cues, the combination of the two. Figure 48 and 49 below shows examples of the cue conditions.

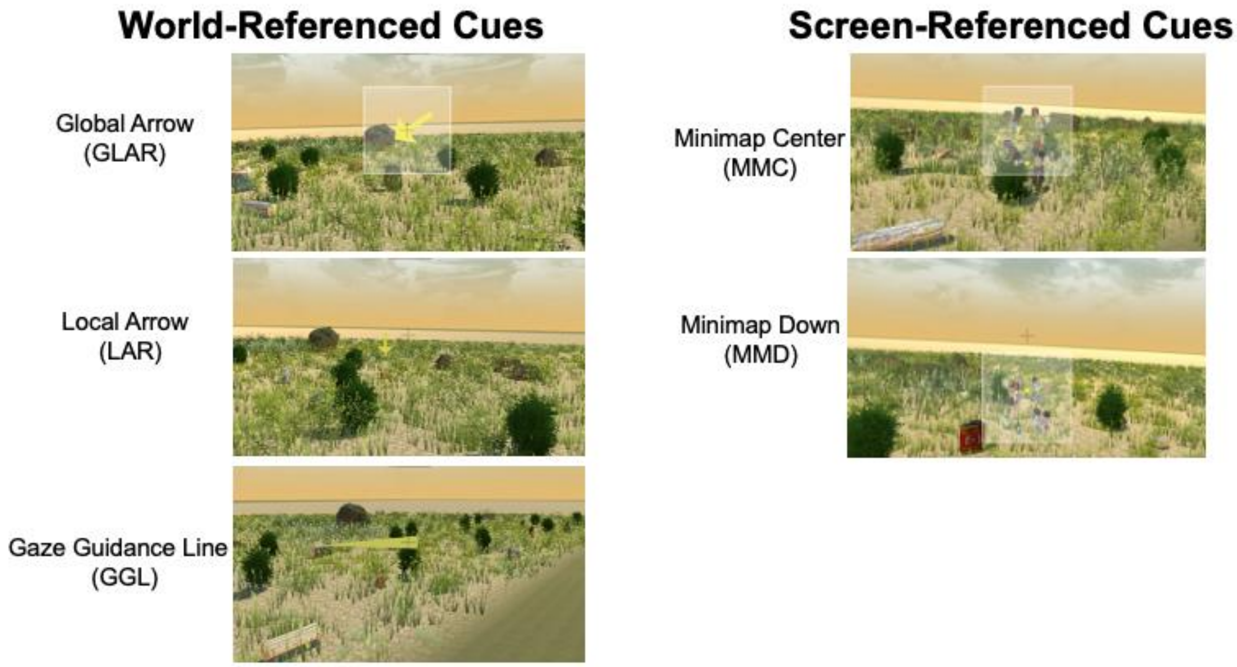

Figure 48. Examples of the single cue conditions when presented in the VR environment. World-reference cues (right) were the global arrow, local arrow, and gaze guidance line. Screen-reference cues were the two minimap conditions: minimap center and minimap 20 degrees down from center.

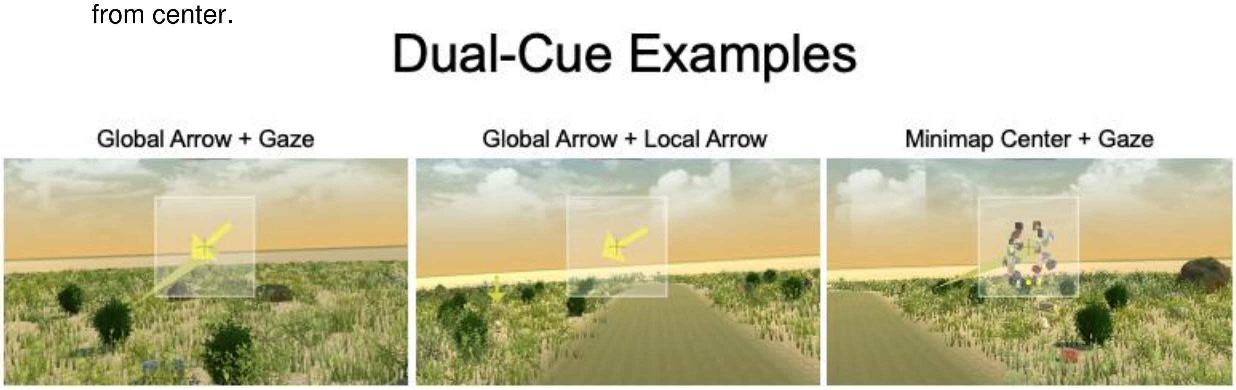

Figure 49. Examples of select dual cue conditions when presented in the VR environment.

Like Experiments 1 and 2, cues were either correct 100% (perfect) or 83% (imperfect) of the time. Imperfect cues errored 17% of the time by cueing the wrong object (i.e., false alarm). The less expected high priority target appeared 17% of the time for each cue condition. On trials where the less expected high priority object appeared, it was never located more than 1 object away from the routine target (i.e., there could be one object between the routine and high priority targets).

Experimental Design. The experiment was a 12 (cue type) x 2 (cue reliability) within-subjects design. Like Experiments 1 and 2, cue type was categorized differently depending on the analysis. The set size for all searches was 22**.**

Before each trial, the Meta Quest 3 displayed text indicating the cue condition for 1.75 s, followed by a 500 ms fixation cross, and then the target for that trial appeared on the display to study for 3 seconds. At the beginning of each trial, participants were required to look forward to ensure a default starting position for each trial. Once the trial started, participants scanned the virtual scene for the routine target. Like Experiment 2, the hitbox appeared around the object participants were oriented towards, signaling to them that the object could be selected.

Participants completed two blocks that were counterbalanced by cue reliability (see Experiment 1 for details). Participants completed all cue conditions, including the no cue, for each reliability block. The cue condition sub-block was counterbalanced within the cue reliability blocks. Within each cue condition sub-block, there was a total of 6 randomized trials for each unique cue condition. The high priority object will appear on 1 of the 6 trials (approximately 17% of the time). For the imperfect cue condition, the cue errored on 1 of the 6 trials (approximately 17% error rate). When the cue errors, it will identify an incorrect target. Each reliability block consisted of 72 trials, for a total of 144 trials in the experiment. Participants also completed 1 practice trial for each cue condition for a total of 12 practice trials before they started the experimental trials. Participants received text feedback (i.e., “correct” or “incorrect”) on the practice trials.

Task. Participants completed a dynamic visual search task in VR. The search task simulated walking along a linear pathway while participants searched for a cued routine target and the uncued and less expected high priority target. Like Experiments 1 and 2, participants were unstructured to find the routine target on every trial but, when the less expected high priority target appears, it should take precedence. The instructions for the were generally the same as the other experiments, except for the specific additions related to the dynamic task (see Appendix C). Critically, the instructions for the perfect and imperfect cue conditions were identical to Experiments 1 and 2. At the beginning of the experiment, participants completed 1 practice trial for each unique cue to familiarize themself with the virtual world and how to complete the search task. Participants began each trial at the same position on the path and could only move in the forward direction as they searched for the target.

Procedure. Before the experiment, participants gave their consent to participate after reading and signing the consent documentation. Participants then read the experimental instructions on a desktop computer where they were also shown example images of the routine and the less expected high priority targets. Before starting the experiment, participants sat in a stationary chair and adjusted their height in the Oculus Meta Quest 3 to ensure that all participants had the same height once they entered the virtual world. Before the experimental trials, participants completed practice trials to familiarize themselves with the virtual environment and controls for the experiment. Using the Meta Quest controllers, participants selected objects by pressing 'A', moved forward along the path using the left controller toggle, and continued to the next trial by pressing 'X.' Once they finished the experiment, they completed the MSSQ motion sickness susceptibility questionnaire, the colorblindness screening, and the same trust in automation and cue questions used in Experiments 1 and 2. The entire experiment lasted approximately 1 hour.

Results

Based on the data from the previous experiments, the number of cues tested in the current experiment was reduced to include only the center minimap (MMC), downward minimap (MMD), global arrow (GLAR), local arrow (LAR), gaze guidance line (GGL), and the no cue condition. Note the difference between the MMC and MMD is that the latter was presented downward on the HMD display by 20 degrees and the former was presented at the center of the immediate FOV of the simulated HMD. Also, note the MMC is the same minimap as the previous experiments (i.e., in the center of the line of sight). This was done to examine the possible detrimental effects of clutter when the minimap was positioned higher in the forward line of sight with an AR-HMD (Warden et al., 2022).

Before conducting any analyses, the data were examined for outliers in the same way as in Experiment 1. Based on these criteria, no participants were deemed as outliers. Four participants only completed half of the experiment and were removed due to incomplete data. All remaining data ($N = 40$) was analyzed using R Studio.

The assumption of normality was tested using the Shapiro-Wilk normality test, and normality was violated for error data despite data transformations (*ps* $< .05$). Therefore, all reported ANOVA analyses for error data will include the Greenhouse-Geisser (GG) correction to account for such violations.

Like Experiments 1 and 2, all response time data includes both correct and incorrect responses. When analyzing the response time data for routine targets, trials in which participants did not make a response or only selected the high priority target were removed from the analysis to ensure that response time pertained only to the routine targets that were selected. Percent error data for the routine targets included all trials.

Descriptive Statistics for Perfect Cue Conditions

Table 5 and 6 below present cue response times and percent errors for the perfect cueing condition only. The longer search time and greater errors in the no-cue condition indicate that the search task was, in general, more difficult than in either Experiments 1 or 2.

Table 5. Experiment 3: Mean response times, measures in seconds, for each single cue and the respective dual-cue combinations. Standard deviations are presented in parentheses.

	Single Cues	**MMC**	**MMD**	**GLAR**	**LAR**	**GGL**
MMC	**11.32 *s*** (*8.36*)				**8.18 *s*** (6.46)	**7.37 *s*** (6.18)
MMD	**10.45 *s*** (7.63)				**6.62 *s*** (5.16)	**6.67 *s*** (6.78)
GLAR	**10.30 *s*** (7.35)				**7.63 *s*** (6.28)	**7.14 *s*** (6.74)
LAR	**7.01 *s*** (6.21)					
GGL	**7.03 *s*** (8.49)					
No-Cue	**13.67 *s*** (10.73)					

Table 6. Experiment 3: Mean percent errors for each single cue and the respective dual-cue combinations. Standard deviations are presented in parentheses.

	Single Cues	**MMC**	**MMD**	**GLAR**	**LAR**	**GGL**
MMC	**16.25%** (37.0)				**7.91%** (27.10)	**6.25%** (24.26)
MMD	**16.25%** (37.0)				**6.67%** (25.0)	**9.17%** (28.92)
GLAR	**16.67%** (37.35)				**7.50%** (26.39)	**8.75%** (28.32)
LAR	**6.67%** (25.0)					
GGL	**5.83%** (23.47)					
No-Cue	**42.50%** (49.5)					

Figure 50 below shows the mean response time (left) and percent error (right) for each perfectly reliable single cue condition. A one-way repeated-measures ANOVA was used to examine the overall effect of cue type for both response time and percent error. Note that the no-cue condition has been removed from this analysis.

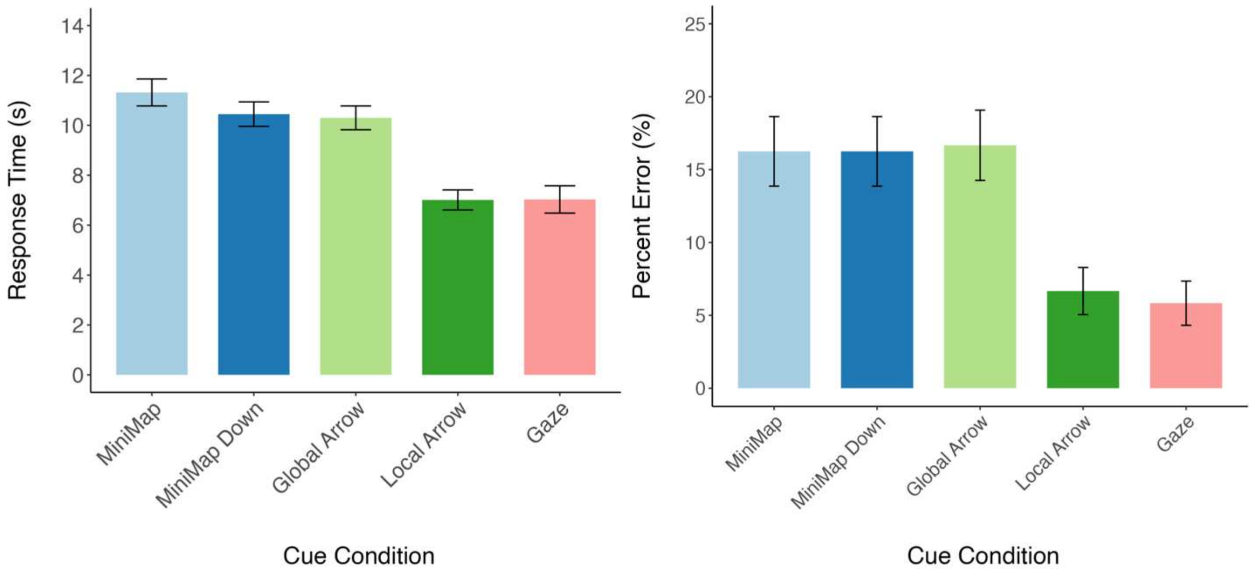

Figure 50. Experiment 3: The mean response time (left) and percent error (right) for all perfectly reliable single cues compared to the no-cue. Error bars represent one standard error of the mean.

The results revealed a significant main effect of cue type on response time, $F(4, 156) = 24.11$, $p < .001$, $\eta_p^2 = 0.38$. The global cues (minimap center, downward minimap and global arrow) showed equivalent search times *(ps* > .099), and were substantially slower than the local arrow ($t(39) = 7.94$, $p < .001$, 95% CI [2.48, 4.17], $d = 0.89$) and gaze guidance line ($t(39) = -5.47$, $p < .001$, 95% CI [-4.52, -2.08], $d = 0.81$), which did not differ from each other ($t(39) = 0.04$, $p = 0.972$, 95% CI [-1.25, 1.30], $d = 0.01$).

The second ANOVA revealed a significant main effect of cue type on percent error, $F(2.25, 87.62) = 4.53$, $p = .01$, $\eta_p^2 = 0.10$. Again, as with response time, the global cues (minimap center, downward minimap, and global arrow) are equivalent (*ps* > .90), and substantially more error prone than the local arrow ($t(39) = 3.02$, $p = .004$, 95% CI [0.03, 0.16], $d = 0.43$) and gaze guidance line ($t(39) = -0.33$, $p = .74$, 95% CI [-0.04, 0.03], $d = 0.03$), which did not significantly differ from each other ($t(39) = -0.33$, $p = .74$, 95% CI [-0.04, 0.03], $d = 0.03$)

Overall, there were significant differences in both response time and percent error between the single, perfect cue conditions. Comparing the results to Experiment 2, the current findings show a greater performance decrement with the global arrow than the local arrow.

Overall Benefits of Cueing

A one-way repeated-measures ANOVA was used to examine the overall effect of cueing. The factor *cue* was collapsed across all possible cue types and cue reliability (i.e., perfect, imperfect). Note that Hypothesis 1 predicted an overall performance benefit of cueing relative to the no-cue condition.

Response Time. The ANOVA results revealed an overall faster response time for the cue (M = 8.98 s) compared to the no cue (M = 13.21 s) condition, $F(1, 39) = 46.16$, $p < .001$, $\eta_p^2 = 0.54$.

Percent Error. A one-way repeated-measures ANOVA examining the effect of cueing on error rate revealed a decrease in error rate in the cued (M = 15.17%) compared to no cue (45.0%) conditions, $F(1, 39) = 105.89$, $p < .001$, $\eta_p^2 = 0.73$.

Confirming Hypothesis 1, search tasks aided by an attention cue enhance search performance, allowing for an improvement of 4.23 s faster and 29.83% more accurate searches compared to an unaided search. These findings translate into a percentage increase of 47.22% in response time and 196.64% in percent error.

Effect of Cue Type and Reliability

The overall effect of reliability was examined via a paired t-test, with the no cue condition excluded, for both response time and percent error.

Response Time. The response time for imperfectly reliable cues (M = 9.80 s) was significantly greater than that for the perfectly reliable cues, (M = 8.16 s; $t(39) = 3.76$, $p = .0006$, 95% CIs [0.76, 2.53], $d = 0.63$).

Percent Error. The percent error for imperfectly reliable cues (M = 20.50%) significantly increased compared to perfectly reliable cues, (M = 9.81%; $t(39) = 7.45$, $p < .001$, 95% CIs [0.08, 0.14], $d = 0.63$.

Hypothesis 1A was supported for both response time and percent error, showing an overall decrement in performance when automation erred.

Effect of Dual-Cueing

The following analysis will examine the joint effects of cue type and reliability for single versus dual cues using the same analyses as Experiments 1 and 2. As a reminder, the greatest benefits of dual cueing are expected to occur with spatial location cues. Therefore, the only dual cues considered were those that combined a local cue (local arrow, gaze guidance line) with a global cue (minimaps, global arrow). The joint effect of dual-cueing and cue reliability was examined using a two-way repeated-measures ANOVA for both response time and percent error. The no cue condition was removed prior to the analysis.

Response Time. The mean response time data are shown in Figure 51. Dual cueing (M = 8.14 s) resulted in significantly faster response times than single cueing (M = 10.0 s), $F(1, 39) = 46.87$, $p < .001$, $\eta_p^2 = 0.55$. Additionally, perfect cues (M = 8.15 s) resulted in significantly faster search times than imperfect cues (M = 9.80 s), $F(1, 39) = 13.65$, $p < .001$, $\eta_p^2 = 0.26$. The interaction was not significant, $F(1, 39) = 0.18$, $p = .675$, $\eta_p^2 < .01$.

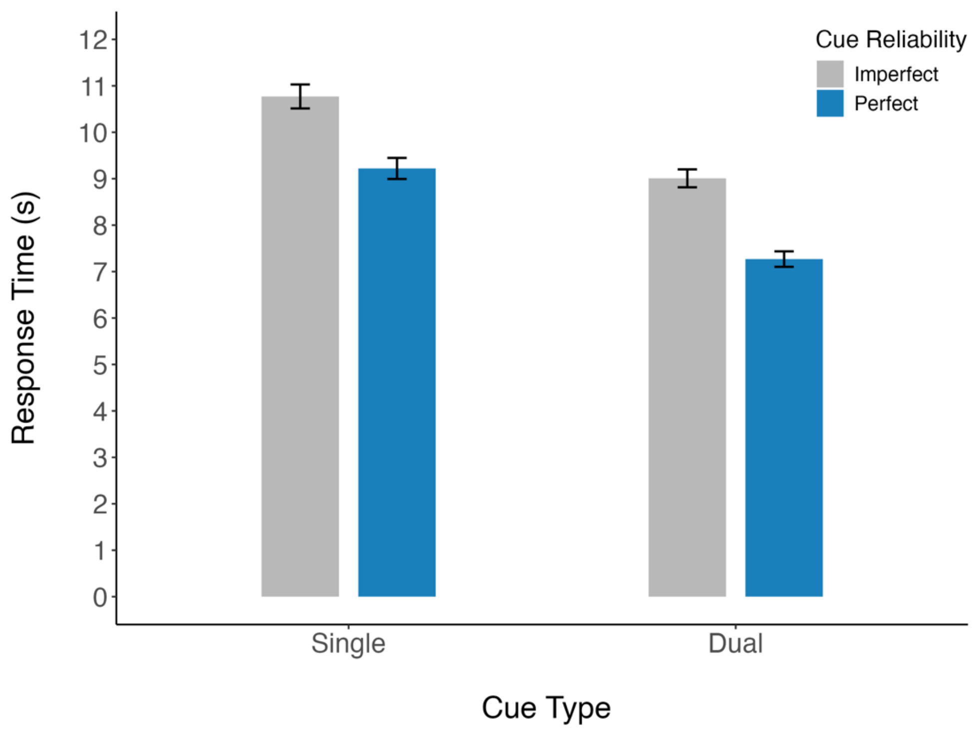

Figure 51. Experiment 3: The mean response time for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

Percent Error. Percent error was analyzed the same way as response time except that percent error was the dependent measure. The mean percent error data are shown in Figure 52. The results revealed that dual cueing (M = 13.23%) significantly reduced errors compared to single cueing (M = 17.50%), $F(1, 39) = 11.20$, $p = .002$, $\eta_p^2 = 0.22$. Additionally, imperfect cues (M = 20.53%) increased errors compared to perfect cues (M = 9.81%), F(1, 39) = 52.78, $p < .001$, $\eta_p^2 = 0.58$. The interaction between cue type and cue reliability was not statistically significant, $F(1, 39) = 0.19$, $p = .664$, $\eta_p^2 < .01$. While not statistically significant, the findings are in the same direction as those found in both Experiments 1 and 2.

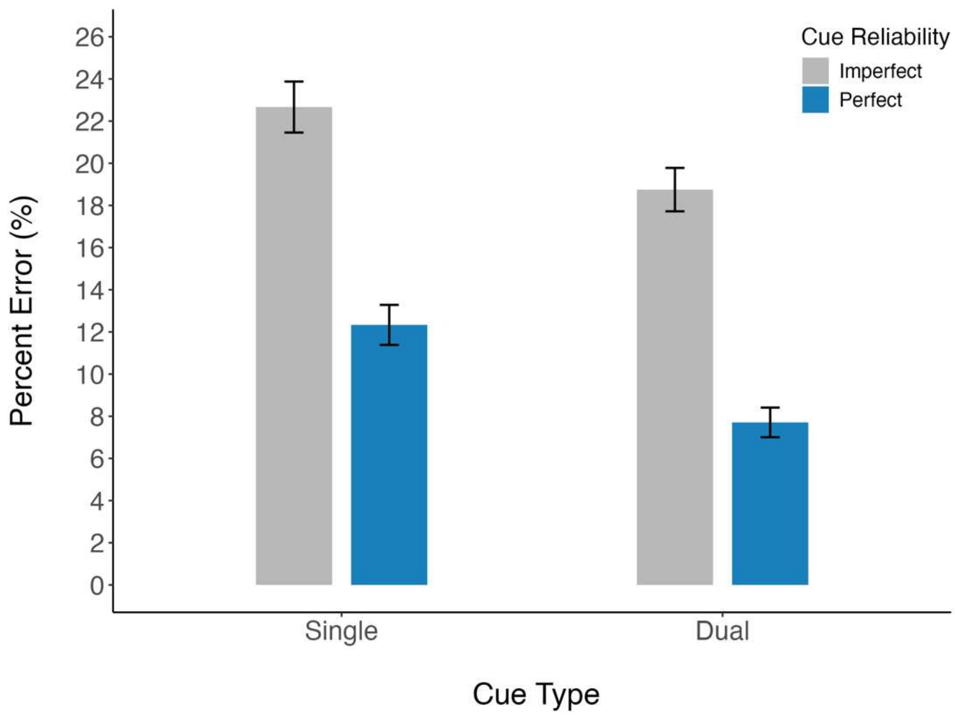

Figure 52. Experiment 3: The mean response time for cue type (single, dual) and cue reliability (gray = imperfect, blue = perfect). Error bars represent one standard error of the mean.

Confirming the findings from both Experiment 1 and 2, dual cueing significantly improved visual search performance by reducing search time and error rates compared to single cues. Additionally, cue reliability continues to play a critical role in search performance, with imperfect cues degrading performance. These findings continue to highlight the importance of dual cueing and cue reliability to optimize search efficiency and accuracy.

The Automation Bias: Correct vs Incorrect Automation Cueing

The automation bias was examined the same way as Experiment 2 for both response time and percent error. In these analyses, the automation bias is examined: that is, the cost in performance when automation errs relative to when automation is correct. Another critical question related to the Lumberjack Hypothesis (H_{1A} and H_{2A}) is how this cost may be modified by how compelling the cue is: specifically, does the cue that works best when automation is correct to show the greatest cost when automation fails?

Response Time. The mean response time for the automation bias is shown in Figure 53 below. A 2 (dual-cueing) x 2 (automation performance) repeated-measures ANOVA was conducted to examine the automation bias for response time as a function of cue type (single versus dual) and automation performance (correct versus incorrect). Replicating the results reported above, dual cueing (M = 9.00 s) was significantly faster than single cueing (M = 10.77 s), $F(1, 39) = 11.87$, $p = .001$ $\eta_p^2 = 0.23$. Additionally, responses were significantly more rapid when automation was correct (M = 9.30 s) compared to incorrect (M = 12.35 s), $F(1, 39) = 43.18$, $p < .001$, $\eta_p^2 = 0.53$). The interaction between cue type and automation performance was not statistically significant, $F(1, 39) = 0.26$, $p = .60$, $\eta_p^2 < .001$.

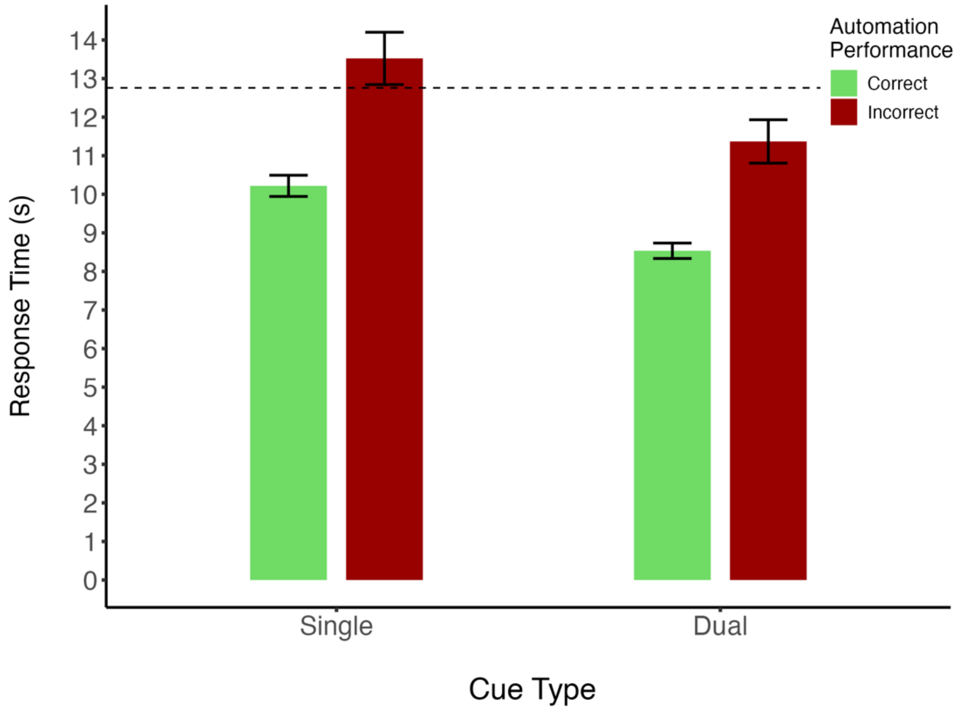

Figure 53. Experiment 3: The mean response time for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no cue condition. Error bars represent one standard error of the mean.

Percent Error. The same analyses were conducted for percent error. The mean percent error data are shown in Figure 54. The main effect of dual cueing was not statistically significant, $F(1, 39) < 0.01$, $p = .994$, $\eta_p^2 < .001$. The main effect of automation performance was significant, $F(1, 39) = 105.42$, $p < .001$, $\eta_p^2 = 0.73$, revealing that error was greater when automation was wrong ($M = 57.73\%$) than when it was right ($M = 13.10\%$), and actually greater than in the unaided condition ($M = 47.50\%$; the dashed line). The interaction between cue type and automation performance was significant, $F(1, 39) = 7.90$, $p = .008$, $\eta_p^2 = 0.17$, signaling that when the cue was correct, accuracy was improved by dual cueing, but when both cues were wrong, dual cueing was more harmful to accuracy than single cueing.

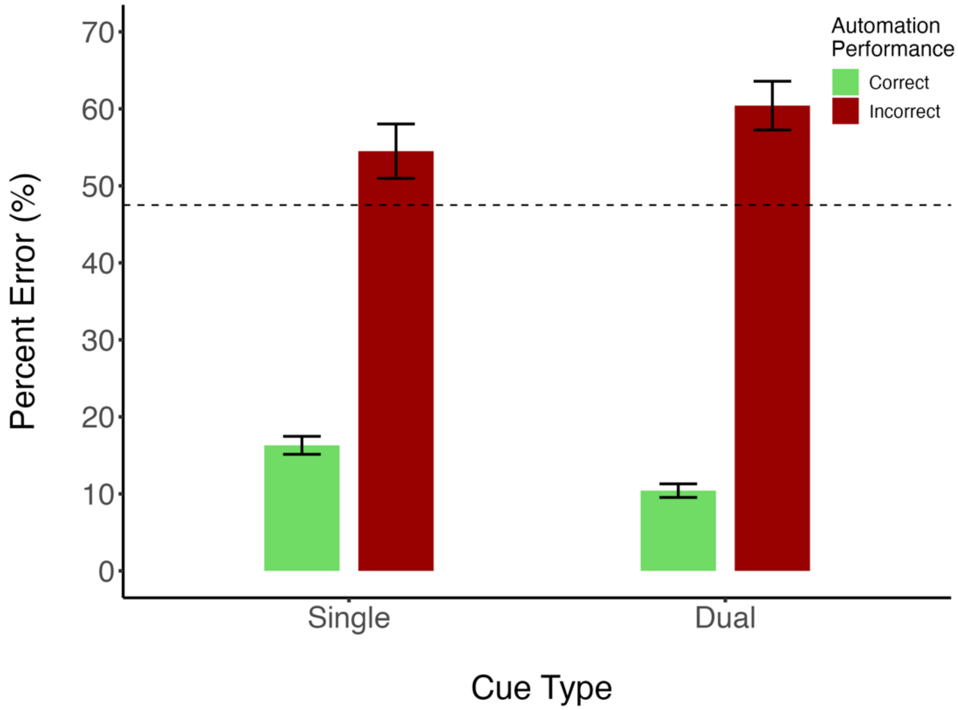

Figure 54. Experiment 3: The mean percent error for cue type (single, dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no cue condition. Error bars represent one standard error of the mean.

As with Experiment 2, the following analysis examined the automation bias for the best and worst cues, specifically the gaze guidance line and the global arrow. The mean response times are shown in Figure 55.

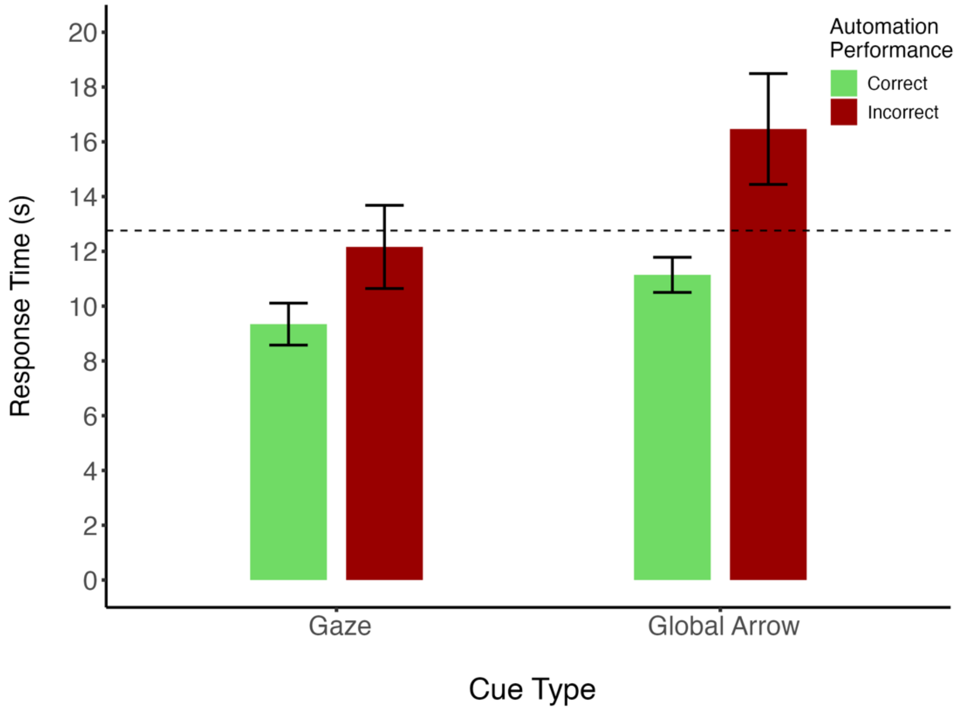

Figure 55. Experiment 3: The mean response time for single cue type (gaze guidance line and global arrow) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no cue condition. Error bars represent one standard error of the mean.

Results from a 2 (cue type) x 2 (automation performance) repeated-measures ANOVA revealed a significant effect of cue type, $F(1, 39) = 6.09$, $p = .02$, $\eta_p^2 = 0.14$. Confirming an overall automation bias, there was a significant main effect of automation performance, $F(1, 39) = 9.26$, $p = .004$, $\eta_p^2 = 0.19$. The interaction between cue type and automation performance was not significant, $F(1, 39) = 1.37$, $p = .392$, $\eta_p^2 = 0.03$.

The following analysis examines differences in automation bias for the gaze guidance line and global arrow (see Figure 56) for error rate – the best and worst single task cues for percent error.

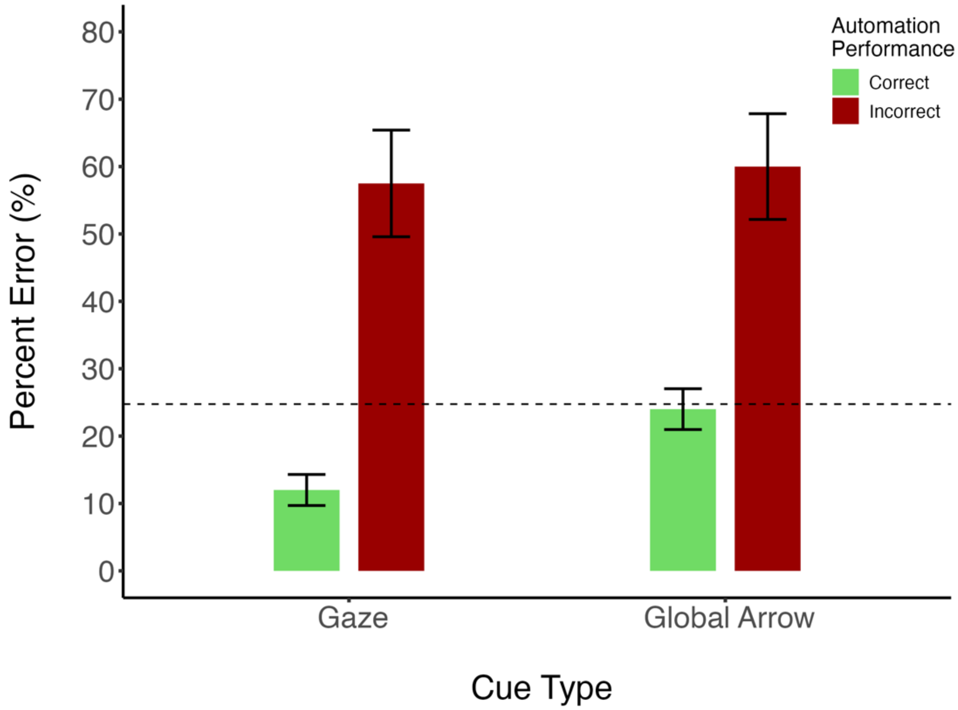

Figure 56. Experiment 3: The mean percent error for single cue types (gaze guidance line and global arrow) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only. The dashed line represents mean response time for the no cue condition. Error bars represent one standard error of the mean.

The ANOVA revealed a significantly higher error rate when the automation erred (M = 58.75%) than when it was correct (M = 18%), $F(1, 39) = 49.36$, $p < .001$, $\eta_p^2 = 0.56$. Neither the effect of cue type, $F(1, 39) = 1.54$, $p = .222$, $\eta_p^2 = 0.04$, nor the interaction, $F(1, 39) = 1.02$, $p = .32$, $\eta_p^2 = 0.03$, were significant. There was no significant difference in the magnitude of the automation bias between the gaze guidance line and global arrow cue, $t(39) = 1.01$, $p = .318$, 95% CI [-0.09, 0.28], $d = 0.20$.

Collectively, these findings highlight that there seems to be a tendency for participants to exhibit the automation bias, as shown when the greater error when automation is wrong versus right. This finding is amplified by the dual-cue, confirming the lumberjack hypothesis (H_{2A}): that is, the more the cue helps accuracy when automation is perfect, the more it hurts accuracy when it errors.

Effect of Cue Precision and Number of Cues: Redundancy Gain

The same analysis as Experiment 2 was used to examine the redundancy gain of the global-local cue combinations compared to either the global or local cues alone. Data were collapsed across cue reliability, and the no cue condition was excluded from the analysis.

Response Time. The mean response time across cue types is presented below in Figure 57. A one-way repeated measure ANOVA revealed a significant effect of cue type on response time, $F(2, 78) = 44.89$, $p < .001$, $\eta_p^2 = 0.54$. Pairwise comparisons (paired t-test) showed that response time was significantly slower with the global cues ($M = 11.14$ s) than the local ($M = 8.28$s; $t(39) = -7.20$, $p < .001$, 95% CIs [-3.66, -2.06], $d = 1.00$) and redundant cues ($M = 8.14$ s; $t(39) = 9.32$, $p < .001$, 95% CIs [2.34, 3.65], $d = 1.14$). There were no significant differences in search time between redundant and local cues, $t(39) = 0.39$, $p = .697$, 95% CIs [-0.57, 0.84], $d = 0.05$, indicating no redundancy gain.

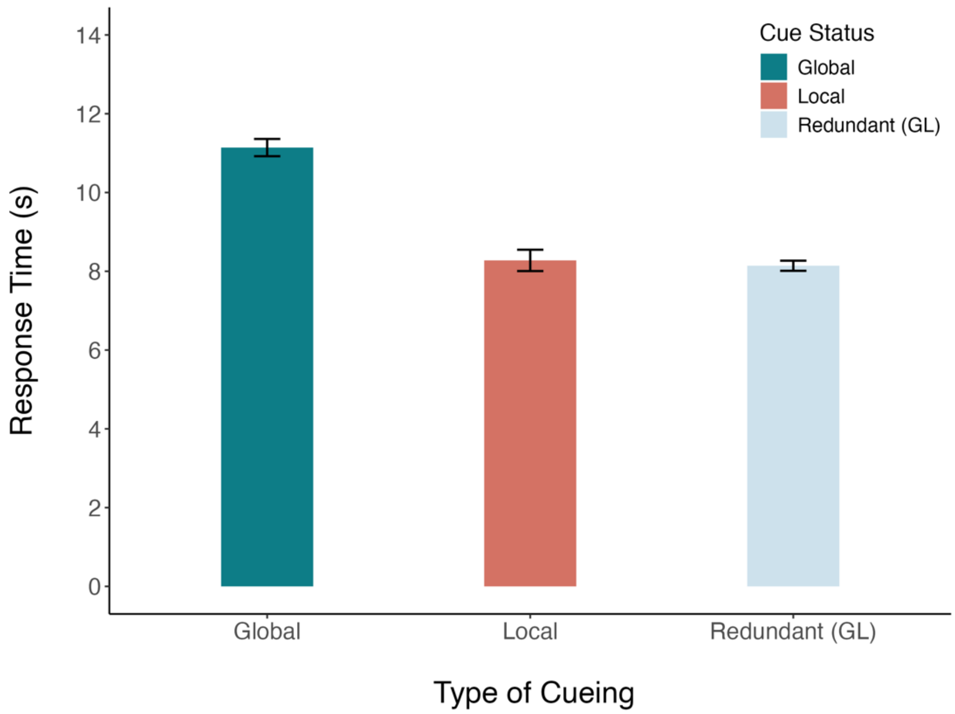

Figure 57. Experiment 3: The mean response time for cue type. Error bars represent one standard error of the mean.

Percent Error. The mean percent error across cue types is presented below in Figure 58. Results revealed a significant effect of cue type on percent error, $F(1.26, 48.99) = 10.43$, $p < .001$, $\eta_p^2 = 0.21$. Pairwise comparisons (paired t-test) showed the same pattern of results as the response time data. Specifically, percent error was significantly higher for global cues ($M = 20.49\%$) compared to the local cues, ($M = 13.02\%$; $t(39) = -3.25$, $p = .002$, 95% CIs [-0.12, -0.03], $d = 0.37$) and redundant cues ($M = 13.23\%$; $t(39) = 3.52$, $p = .001$, 95% CIs [0.03, 0.11], $d = 0.36$). Percent error was equivalent between local and redundant cues, $t(39) = -0.22$, $p = .82$, 95% CIs [-0.02, 0.02], $d = 0.01$. Like response time, there was no true redundancy gain.

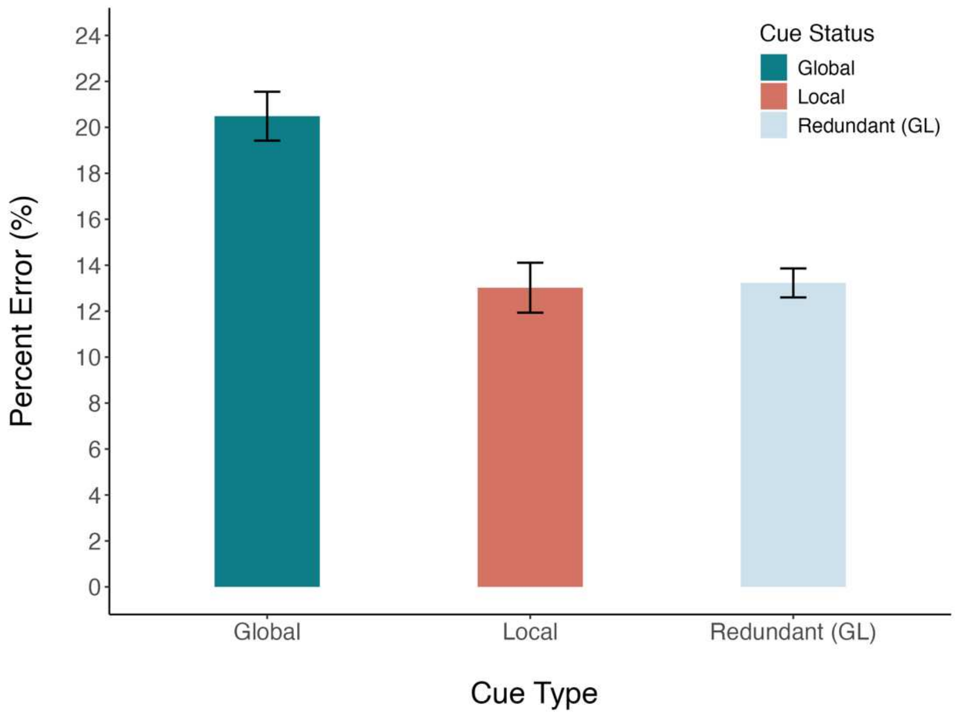

Figure 58. Experiment 3: The mean percent error for cue type. Error bars represent one standard error of the mean.

Collectively, the pattern of results for response time and percent error show that both local and redundant (global + local) cues improve overall search efficiency and accuracy, but there is no additional performance benefit of global cues. While a redundancy gain for response time was found in Experiment 2 with a true AR-HMD, no such gain was found for a simulated HMD in virtual reality. These findings highlight that cueing to local information or combining local cues with global information are similarly optimal for improving search performance.

Attentional Tunneling

The effect of attentional tunneling was examined in the same way as Experiment 2. Only trials where a less expected high priority target object appeared were included. For response time data, only trials where they actually selected a high priority object were included because these are the only trials for which there is response time data for selecting that object. All data are collapsed across cue reliability and cue type.

Response Time. The first analysis examined whether the cue status, either cued or uncued, of the routine target impacted the searchers' ability to find the less expected high priority target occurring in the same scene. The mean response times are shown in Figure 59.

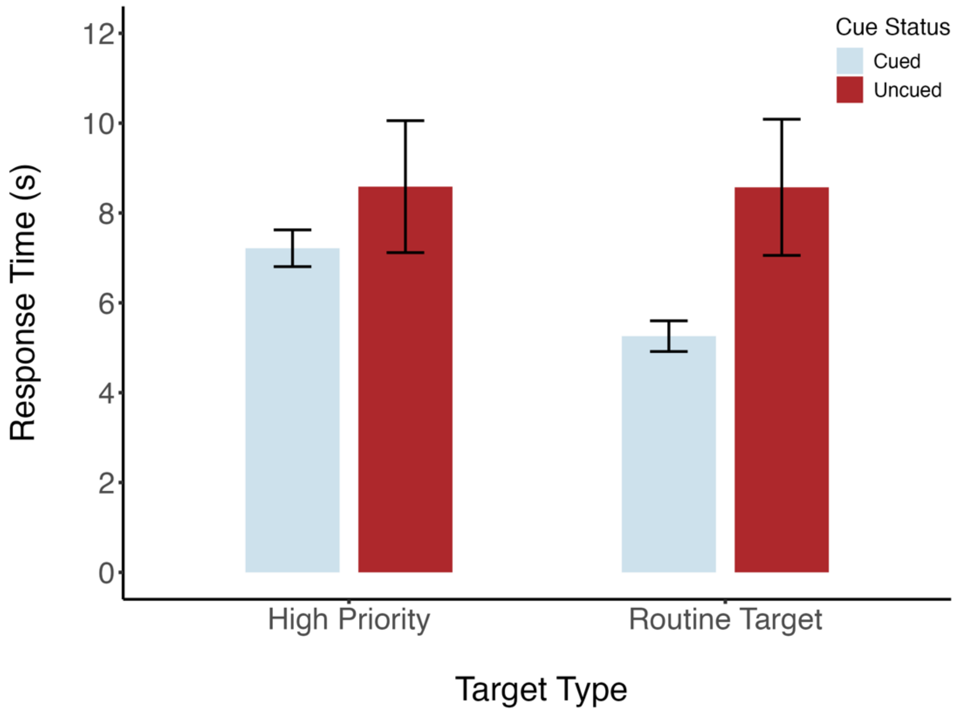

Figure 59. Experiment 3: Mean response time for the less expected high priority target (left) or the routine target (right) as a function of whether the routine target was cued (collapsed over all cue types; light blue) or uncued (red).

A total of 14 participants were removed from the analysis because they failed to find the high priority target on specific trials resulting in missing response time data (note that these participants are included in the error data since they missed the high priority target on these trials). A two-way repeated measures ANOVA revealed a significant main effect of cue status on response times, $F(1, 17) = 8.12$, $p = .01$, $\eta_p^2 = 0.32$, showing higher response times for both targets when the routine target was uncued. Neither the effect of target type ($F(1, 17) = 0.50$, $p = .491$, $\eta_p^2 = 0.03$) nor the interaction between target type and cue status ($F(1, 17) = 0.84$, $p = .373$, $\eta_p^2 = 0.05$) were significant.

Figure 60 shows the mean percent error for high priority and routine targets as a function of cue status (cued, uncued).

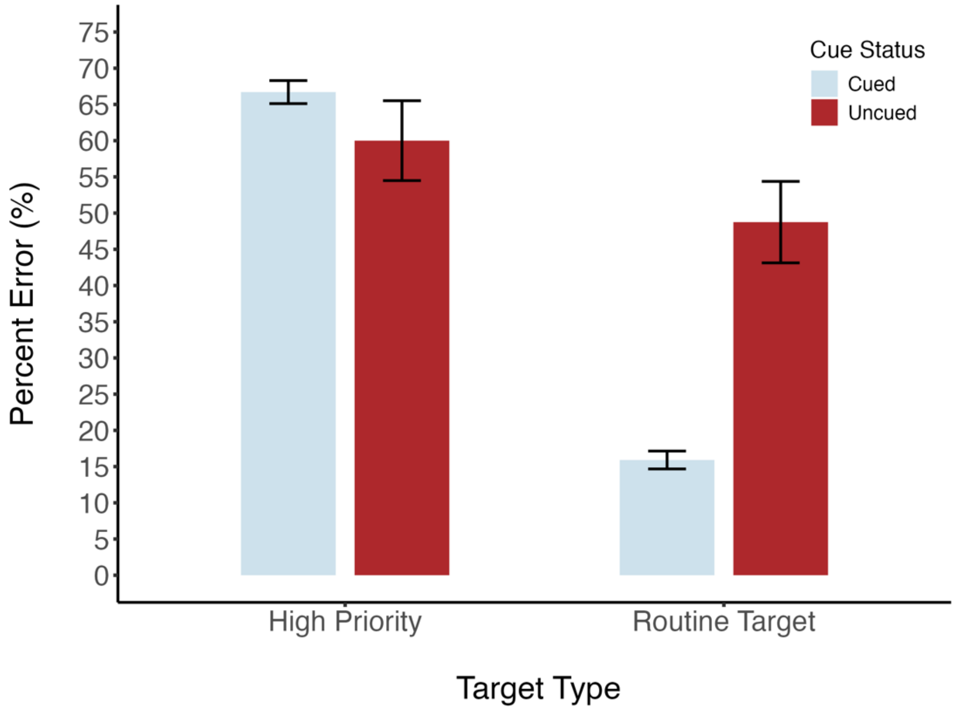

Figure 60. Experiment 3: Mean percent error for the less expected high priority target (left) or the routine target (right) as a function of whether the routine target was cued (collapsed over all cue types; light blue) or uncued (red).

The ANOVA revealed a significant main effect of target type ($F(1, 39) = 30.34$, $p < .001$, $\eta_p^2 = 0.44$) and cue status ($F(1, 39) = 14.24$, $p < .001$, $\eta_p^2 = 0.28$). These main effects are better interpreted in the context of the significant interaction, $F(1, 39) = 24.56$, $p < .001$, $\eta_p^2 = 0.39$. Pairwise comparisons show, that for routine targets, there is a large reduction of error when those targets are cued ($t(39) = -5.64$, $p < .001$, 95% CI [-0.45, -0.21], $d = 1.03$. However, the presence of the cue for the routine target had no significant impact on the detection of the high priority targets that were present in the same visual field, $t(39) = 1.43$, $p = .16$, 95% CI [-0.03, 0.17], $d = 0.19$.

Figure 61 shows the mean response time for high priority and routine targets as a function of dual cueing.

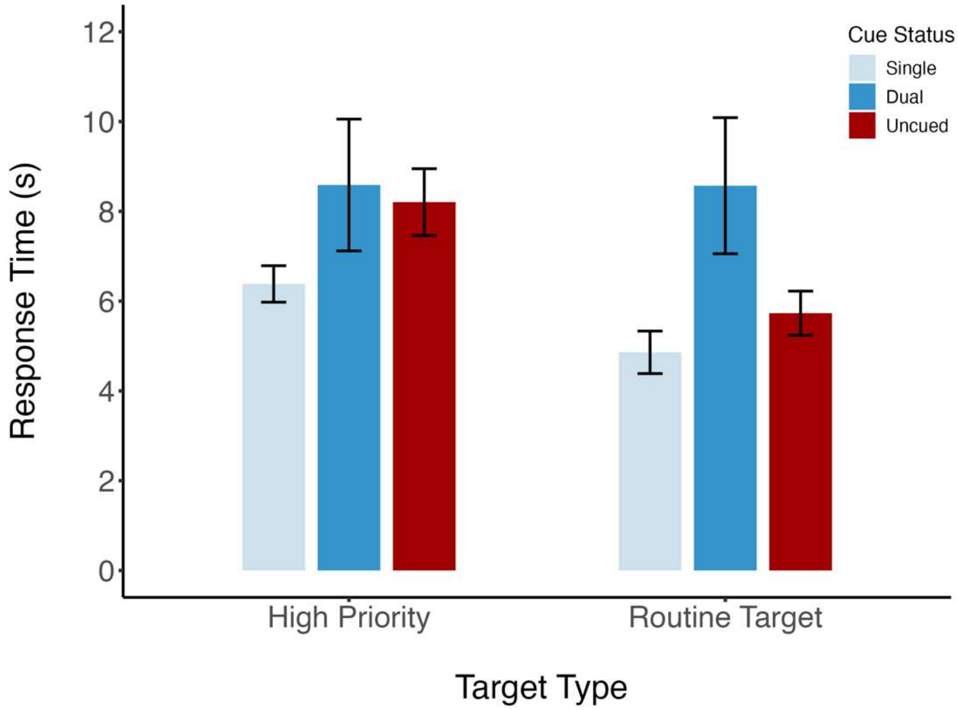

Figure 61. Experiment 3: Mean response time for the less expected high priority target (left) or the routine target (right) as a function of whether the routine target was cued with a single (light blue) or dual (dark blue) cue, or uncued (red).

The ANOVA revealed a main effect of cue status, $F(1.26, 21.47) = 6.56$, $p = .013$, $\eta_p^2 = 0.28$. Neither the main effect of target type ($F(1, 17) = 1.81$, $p = .196$, $\eta_p^2 = 0.1$) nor the interaction ($F(1.27, 21.64) = 0.70$, $p = .446$, 0.04) were significant.

Figure 62 shows the mean percent error for high priority and routine targets as a function of dual cueing.

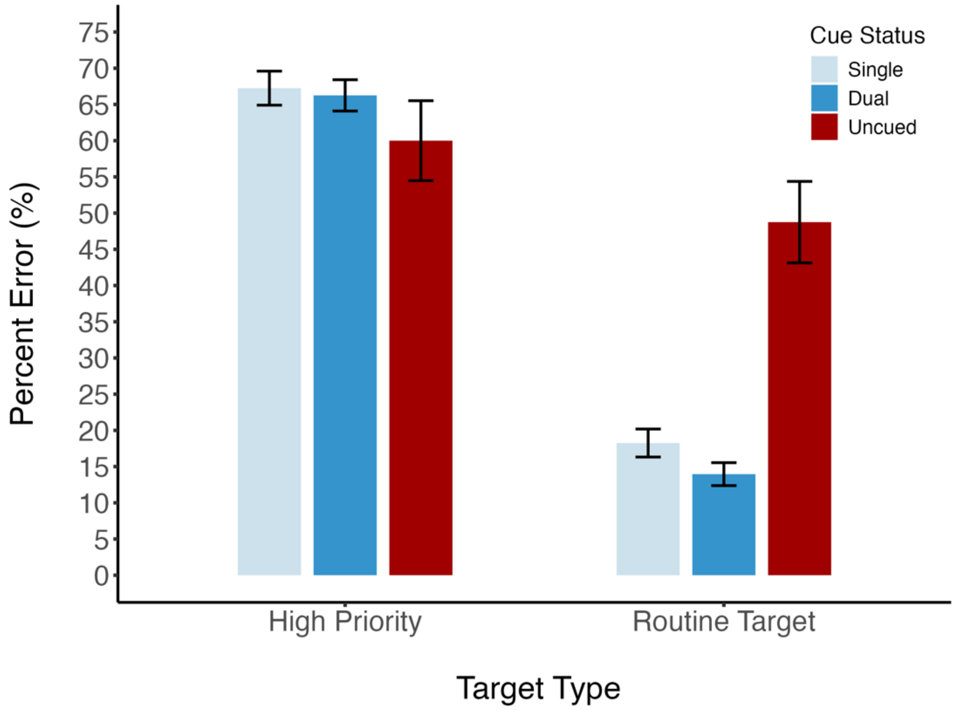

Figure 62. Experiment 3: Mean percent error for the less expected high priority target (left) or the routine target (right) as a function of whether the routine target was cued with a single (light blue) or dual (dark blue) cue, or uncued (red).

The ANOVA revealed a significant main effect of target type ($F(1, 39) = 51.26$, $p < .001$, $\eta_p^2 = 0.57$) and cue status ($F(1.33, 51.96) = 12.23$, $p < .001$, $\eta_p^2 = 0.24$). These main effects are better interpreted in the context of the significant interaction, $F(1.31, 51.18) = 20.68$, $p < .001$, $\eta_p^2 = 0.34$.

Overall, the findings above show that cue status, whether the routine target was cued or uncued, there is a general trend for response time to decrease in the presence of cues. Additionally, while cues decreased errors for routine targets, they did not significantly impact the detection accuracy for the less expected high priority target. However, the pattern of results are in the general direction of increasing errors when the routine target is cued. The lack of significance may be due to the fact that there were so few trials where the expected high priority target appeared (i.e., low statistical power).

Effect of Cue Display Imagery: World vs Screen-Referenced Coordinates

To examine the effects of cue display imagery, cue types were categorized based on whether the cue was in world-referenced (conformal) coordinates, meaning that the cue directly links to or overlays the object in the real world, or whether the cue is in screen-referenced (nonconformal) coordinates, meaning the position of the cue is linked to the display of the device. The following analysis presents cue display imagery for only single cues: global arrow (GLAR), gaze guidance line (GGL), local arrow (LAR), minimap center (MMC), and minimap down (MMD). Note that the no-cue condition was excluded from the analysis. The mean response time data are below in Figure 63.

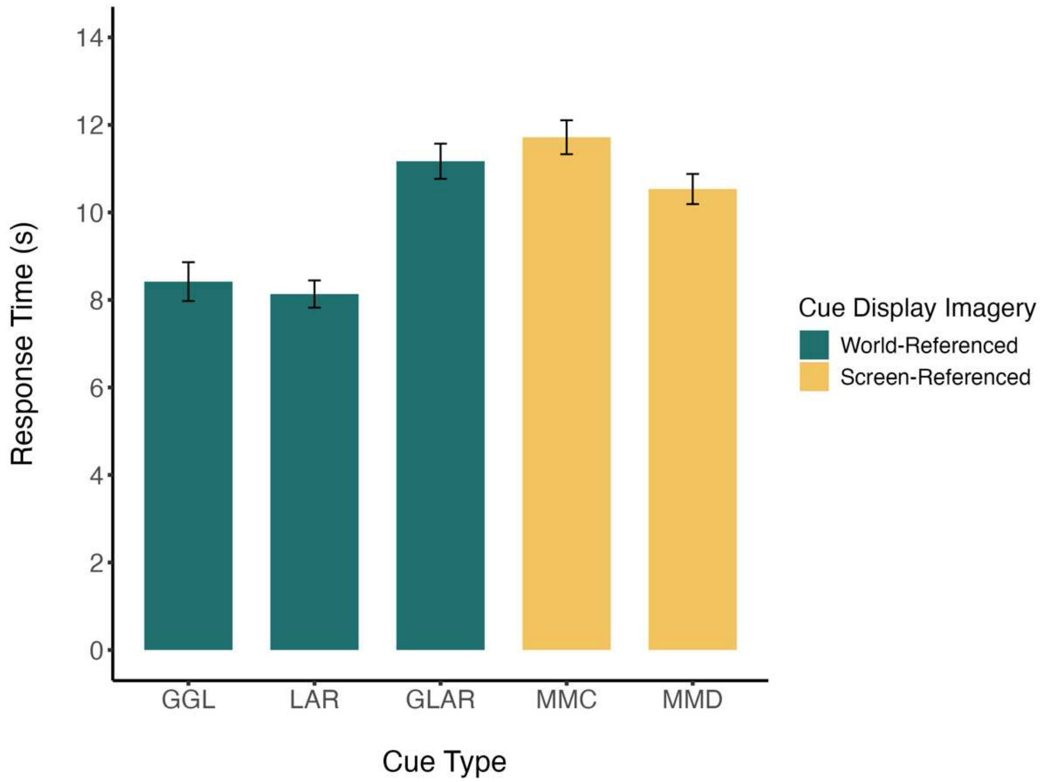

Figure 63. Experiment 3: Mean response time as a function of single cues that are either world-referenced (teal) or screen-referenced (yellow). Error bars represent one standard error of the mean.

A one-way repeated measures ANOVA was used to examine the overall effect of cue type. Results revealed a significant main effect of cue conformality, $F(3.56, 138.87) = 16.61$, $p < .001$, $\eta_p^2 = 0.30$. Pairwise comparisons show that the two local, world-referenced cues led to the fastest response times compared to all other cues ($ps < .0015$). The global, world-referenced cue (GLAR; $M = 11.17$ s) was not significantly faster than the global, screen-referenced MMC ($M = 11.72$ s) and MMD cues ($M = 10.53$ s; $ps > .22$). Of the screen-referenced global cues, the downward minimap (MMD) was significantly faster than the center minimap (MMC), $t(39)$, 2.31, $p = .03$, 95% CI [0.15, 2.21], $d = 0.36$, indicating that the lower placement helps mitigate costs of clutter imposed by the minimap when positioned at the center of the user's field of view on the display.

A one-way ANOVA was used to examine the overall effect of cue type for single cues only. The mean percent error is shown in Figure 64.

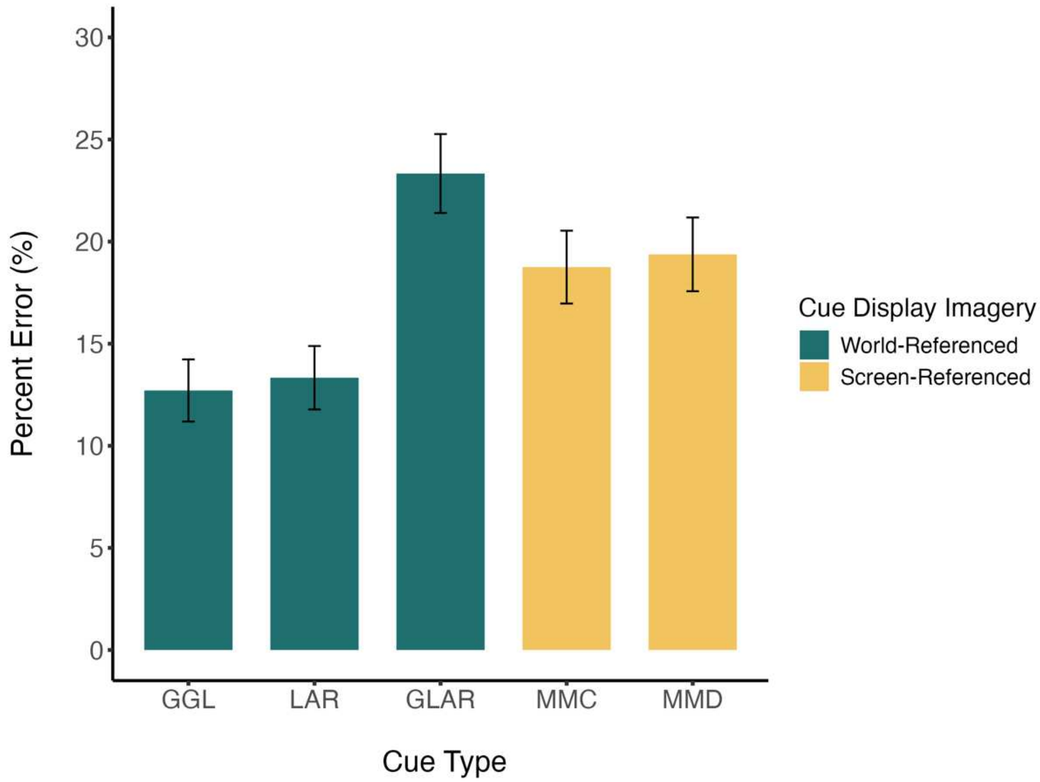

Figure 64. Experiment 3: Mean percent error as a function of single cues that are either world-referenced (teal) or screen-referenced (yellow). Error bars represent one standard error of the mean.

Results revealed a significant main effect of cue type on percent error, $F(4.79, 186.83) = 27.93$, $p < .001$, $\eta_p^2 = 0.42$. Pairwise comparisons indicated that the local, world-reference arrow cue significantly lowered error rates compared to the global, world-referenced arrow cue ($t(39) = 3.92$, $p < .001$, 95% CI [0.05, 0.15], $d = 0.50$) and the downward minimap cue ($t(39) = -2.06$, $p = 0.047$, 95% CI [-0.12, -0.001], $d = 0.29$), but no differences in error rate were found compared to the central minimap ($t(39)$ -1.57, $p = .13$, 95% CI [-0.12, 0.02], $d = 0.23$). Similarly, the local, world-reference gaze cue significantly reduced error rates compared to the downward minimap ($t(39) = -2.58$, $p = .014$, 95% CI [-0.12, -0.01], $d = 0.32$) but not the central minimap ($t(39) = -1.92$, $p = .062$, 95% CI [-0.12, 0.003], $d = 0.26$).

While these local cues did not differ from the central minimap, the difference in percent error is in the expected direction (i.e., greater for the screen-referenced cue than the world-reference cue). Surprisingly, no significant differences were found between the global, world-referenced arrow cue and the global, screen-referenced minimap cues (*ps* > .10), indicating these cues supported the worst performance.

Next, a one-way repeated-measures ANOVA was used to examine the effect of all levels of cue imagery (i.e., combinations of imagery). Note that the no cue was removed from the analysis. The mean response times are shown in Figure 65.

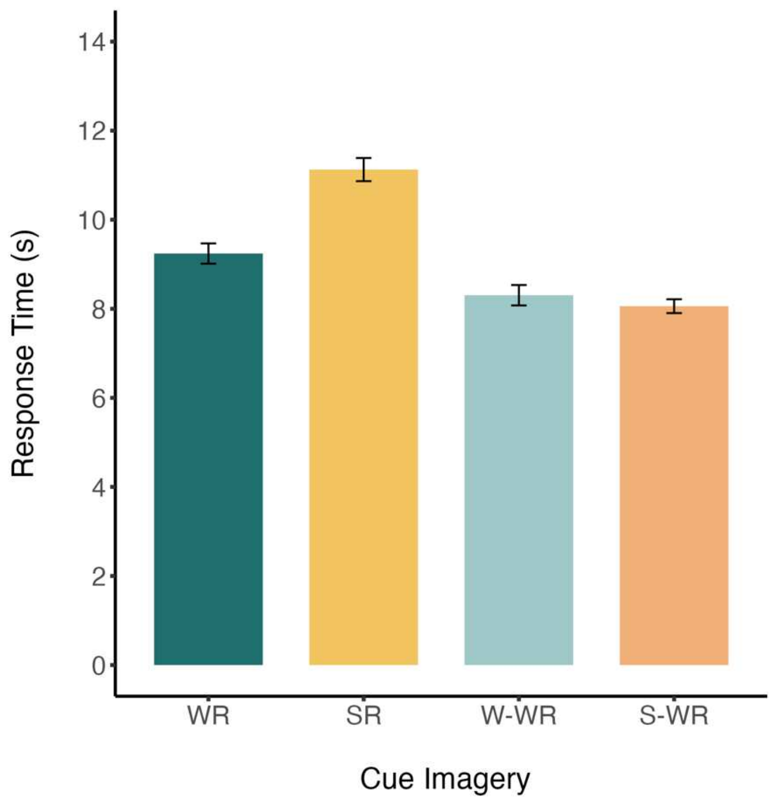

Figure 65. Experiment 3: Mean response time as a function of cue imagery for world-referenced (WR), screen-referenced (SR), world-world referenced (W-WR) and screen-world referenced (S-WR) cues. Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue imagery, $F(2.48, 96.63) = 38.26$, $p < .001$, $\eta_p^2 = 0.50$. Pairwise comparisons show that dual world-referenced cues (W-WR; $M = 8.30$ s) result in lower response times than a single world-referenced ($M = 9.24$ s; $t(39) = 3.72$, $p = .0006$, 95% CI [0.42, 1.44], $d = 0.36$) and single screen-referenced cues ($M = 11.13$ s; $t(39) = 7.66$, $p < .001$, 95% CI [2.07, 3.56], d =

1.09). But there were no significant differences in response time between the dual world-reference and dual screen-reference cues, (*M* = 8.06 s; *t*(39) = -0.91, *p* = .37, 95% CI [-0.80, 0.31], *d* = 0.11.

Next, a one-way repeated-measures ANOVA was used to examine the effect of all levels of cue imagery (i.e., single and dual combinations of imagery). Note that the no-cue was removed from the analysis. See Figure 66 below for percent error data.

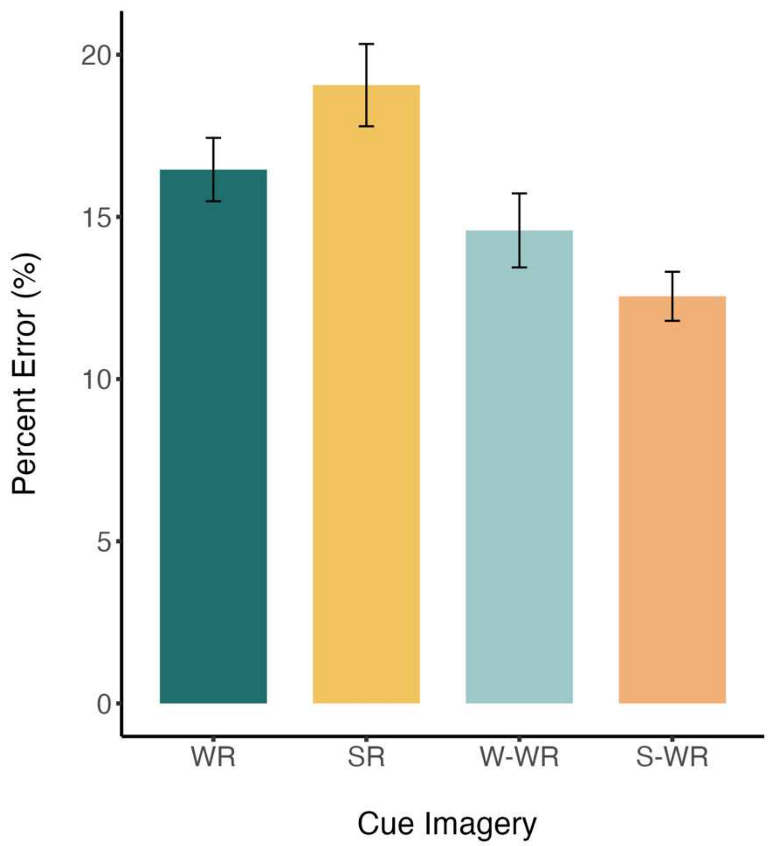

Figure 66. Experiment 3: Mean percent error as a function of cue imagery for world-referenced (WR), screen-referenced (SR), world-world referenced (W-WR) and screen-world referenced (S-WR) cues. Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue imagery, $F(1.61, 62.84) = 3.66$, $p = .0145$, $\eta_p^2 = 0.09$. Pairwise comparisons show percent error was less for the dual cue using screen-world referenced imagery than the single screen-referenced imagery ($t(39) = 2.64$, $p = .012$, 95% CI [0.02, 0.11], $d = 0.29$), but all other comparisons were not statistically significant (*ps* > .12).

In summary, these findings highlight the influence that display imagery has on search performance. Local, world-reference cues consistently outperformed all other types, as shown by the faster search times and lower error rates than global cues and screen-referenced cues. More critically, these findings further confirm the benefit of dual-cueing. Dual-cues were more effective at reducing search time and error rate than either screen or world-referenced single cues alone. Another key finding is that the downward minimap, which helps minimize the amount of overlay clutter occluding the real-world scene beyond the simulated HMD, led to better performance than the minimap positioned at the central field of view. This finding highlights the importance of cue placement on performance. Overall, conformal cues and strategic placement of information can result in better performance.

Field of View Analyses

The field of view analyses were conducted the same way as Experiment 2. The analysis was conducted using a 2 (FOV) x 2 (cue precision) repeated-measures ANOVA. The mean response time results are shown in Figure 67 below.

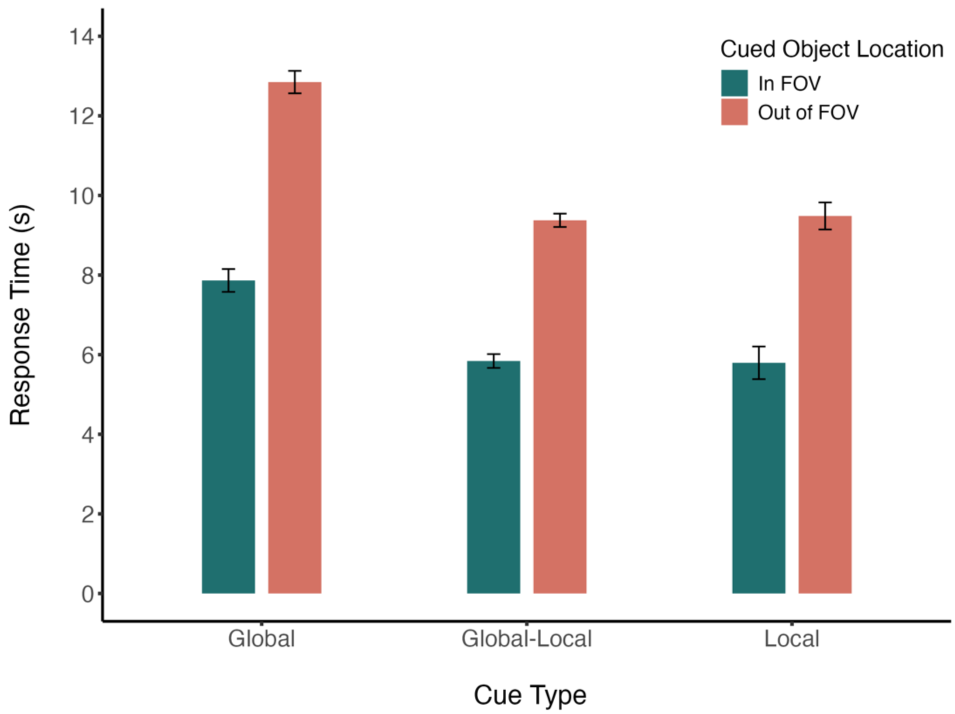

Figure 67. Experiment 3: Mean response time data for cue type (global, local, or global-local) as a function of whether the cues were in (green) or out (red) of the immediate field of view of the device. Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue precision, $F(1.85, 72.13) = 41.16$, $p < .001$, $\eta_p^2 = 0.51$, showing that cues with local spatial information result in the fastest search times. There was also a main effect of FOV, $F(1, 39) = 146.97$, $p < .001$, $\eta_p^2 = 0.79$. As expected, response times are faster when the cued target is in the immediate FOV. The interaction was not significant, $F(1.78, 69.38) = 1.51$, $p = .101$, $\eta_p^2 = 0.04$.

Percent error was analyzed the same way as response time data. The data are presented in Figure 68 below.

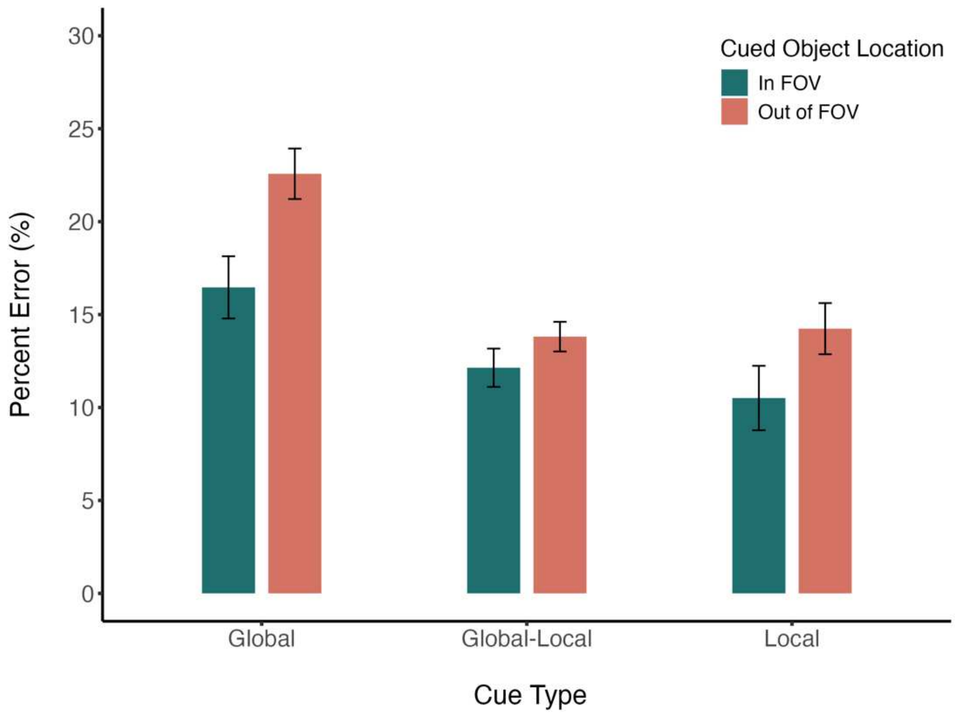

Figure 68. Experiment 3: Mean percent error data for cue type (global, local, or global-local) as a function of whether the cues were in (green) or out (red) of the immediate field of view of the device. Error bars represent one standard error of the mean.

The ANOVA revealed a significant effect of cue precision, $F(1.94, 75.79) = 49.30$, $p < .001$, $\eta_p^2 = 0.56$. There was also a main effect of FOV, $F(1, 39) = 25.62$, $p < .001$, $\eta_p^2 = 0.40$, showing fewer errors when the cued target falls within the immediate FOV. These effects are better explained in the context of the significant interaction, $F(1.76, 68.71) = 14.45$, $p < .001$, $\eta_p^2 = 0.27$. There is a larger cost when targets are outside the immediate field of view, but this cost is amplified in the global cue condition compared to the global-local ($t(39) = -4.62$, $p < .001$) and local ($t(39) = 4.07$, $p < .001$).

Collectively, the location of the cued target object relative to the searcher's immediate FOV significantly impacts both search time and accuracy, showing better performance when the cued target is within the FOV. Additionally, cues incorporating local spatial information substantially reduce search time and error rate relative to the global cue alone, which shows a large cost when targets are outside the FOV. While both single and dual cues featuring a local cue improve search performance, the substantial benefits of the local cues suggest they play a critical role. It may be the case that the virtual environment search scene allowed participants to exploit peripheral vision, and, therefore, allowed them to see the local cues.

Discussion

Replicating the findings from Experiments 1 and 2, cueing improved search performance when compared to an unaided search. The magnitude of the cueing benefit for response time was greater than found in the desktop static visual search task of Experiment 1($t(85) = 3.84$, $p < .001$, $d = 0.80$) but the cueing benefit was equivalent to that found in the AR-HMD static visual search task ($t(71) = 0.21$ $p = .42$, $d = 0.05$). For percent error, the cueing benefit in the VR dynamic visual search task was greater than both the desktop search task ($t(85) = 6.49$, $p < .001$, $d = 1.37$) and the AR-HMD static search task ($t(71) = 2.90$, $p = .002$, $d = 0.80$). Also confirming H_2 and replicating findings from Experiments 1 and 2, dual-cueing led to superior search performance compared to single cueing for both response time ($d = 0.99$) and percent error ($d = 0.53$). However, this benefit of dual-cueing was notably smaller than that found with the AR-HMD (Experiment 2). It may be that participants could use their peripheral vision in the virtual environment because it was not constrained by an actual HMD.

Confirming H_{1A}, performance was worse with imperfect cueing aids. That response time increased with imperfect automation might suggest that people were not entirely blindly following the automation's recommendation and instead opted to spend more time searching. However, given the notable increase in error rate, and considering the general complexity of the search task in VR, there is enough evidence to suggest some degree of automation bias. Unlike the previous experiments, the pattern of results for the dual vs single cue is suggestive of the Lumberjack Effect (Sebok & Wickens, 2017), where the most effective cue, in this case, dual cues, when automation is perfect results in worse performance when automation errors.

Aligning with the previous experiments, evidence for attentional tunneling was weak. However, the pattern of results aligns with the phenomenon. More specifically, the error rate for the less expected high priority target is always greater when the routine target is cued compared to uncued, suggesting some degree of attentional tunneling.

Replicating the finding from Experiment 2 and supporting H4A, single and dual-cues using world-referenced coordinates led to improved performance, suggesting a benefit of cues that link to the spatial location of real-world information. That dual-cues were the most effective further supports duel-cueing systems. Additionally, this experiment found that placing the minimap 20 degrees downward from the forward line of sight improved performance by reducing the amount of overlay clutter.

Exploratory Between-Experiment Analyses: Automation Bias and Attentional Tunneling

Across all three experiments there were two non-significant but consistent trends: (1) that dual-cueing might show a greater automation bias than single cueing and (2) that cueing a routine target will result in an automation-based attentional tunneling for an uncued high priority target. To further explore these trends, data from the three experiments was combined.

The first analysis examined the error data from the imperfect cue conditions only to assess whether dual-cueing, which showed better performance overall, led to a greater degree of automation bias than single cueing. To analyze this effect a mixed ANOVA was used where the between-subjects factor was the experiment and the within-subjects factors were the cue type and the automation performance. See Figure 69 below for the mean percent errors as a function of automation performance.

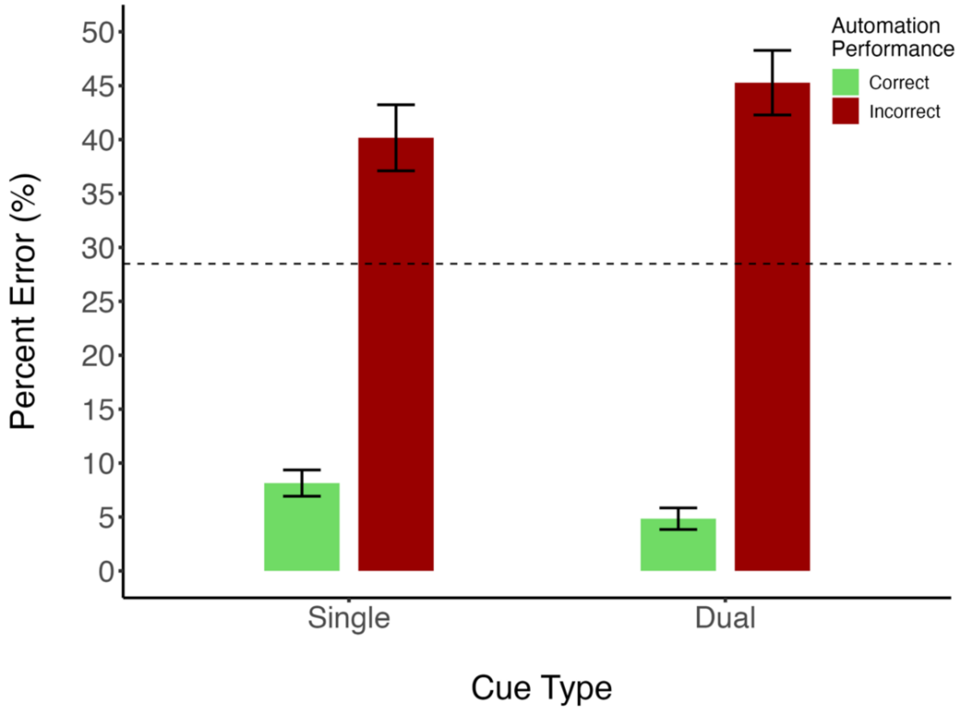

Figure 69. The mean response time for cue type (single versus dual) and automation performance (green = correct, red = incorrect) for the imperfect cue condition only, averaged across all three experiments. The dashed line represents mean response time for the average no-cue condition across all three experiments. Error bars represent one standard error of the mean.

The critical finding lies in the interaction between cue type (single versus dual) and automation performance (correct versus incorrect), which was statistically significant, $F(1, 117) = 12.35$, $p < .001$, $\eta_p^2 = 0.1$. Pairwise comparisons (paired t-test) show that the dual cue led to significantly more errors than the single cue when the automation was incorrect, $t(119) = -2.31$, $p = .02$, 95% CIs [0.09, 0.01], although this effect is considered small as indicated by a Cohen's d of 0.15.

The next analysis examined whether there was significant automation-based attentional tunneling observed for the uncued, less expected high priority target when the routine target was cued versus uncued. The first mixed ANOVA examined this effect collapsed across all cue types (see Figure 70 below).

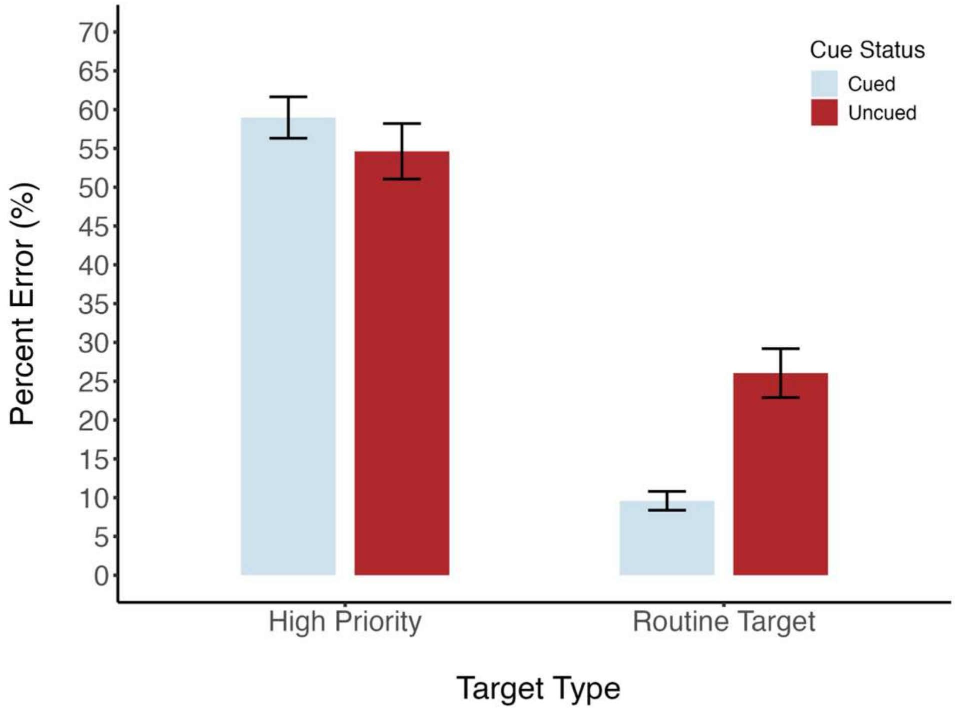

Figure 70. Mean percent error from all three experiments for the less expected high priority target (left) or the routine target (right) as a function of whether the routine target was cued (collapsed over all cue types; light blue) or uncued (red).

A higher percent error for the high priority target when the routine target was cued versus uncued would be indicative of automation-based attention tunneling. A paired t-test revealed no significant differences in the error rate for the high priority target when the routine target was cued versus uncued, $t(118) = 1.50$, $p = .14$, 95% CIs [-0.01, 0.10], $d = 0.12$. The same analysis was conducted for the dual versus single cues. Again, results failed to find a significant difference between either the dual ($t(118) = 1.54$, $p = 0.13$, 95% CIs [-0.01, 0.11], $d = 0.13$) or single cues ($t(118) = 1.31$, $p = 0.19$, 95% CIs [-0.02, 0.10], $d = 0.11$) compared to when the routine target was uncued. While the pattern of results for both comparisons is in the expected direction (i.e., higher errors for the high priority when the routine target is cued versus not cued), the finding again suggests no significant automation-based attentional tunneling was observed after pooling the data together from all three experiments. This is the case for cueing, in general, and dual cueing (i.e., dual cueing does not impose a greater degree of attentional tunneling than single cueing).

GENERAL DISCUSSION

The goal of the three experiments was to examine cue effectiveness, generally, and for an imperfect dual-cueing system that may err on occasion. Cue effectiveness was examined in the context of a visual search paradigm where participants searched for a routine target and a less expected high priority target. The visual search task increased in its complexity, scale, and realism, starting with a static visual search on a 2D desktop monitor (Experiment 1) to a static search presented with an AR-HMD with a larger 2D search scene (Experiment 2) to an even more realistic simulated dynamic search task with both display and scene presented in virtual reality (Experiment 3). Visual search performance (i.e., response time and accuracy) during an aided and unaided search were measured in all three experiments. Table 7 below summarizes the main effects examined across the three experiments and provides the basis for extracting general trends from the many specific results described previously.

Table 7. The table below presents the average Cohen's d effect sizes for each effect investigated in all three experiments. Only the main effects were used when computing the average effect sizes. HP represents the High Priority Target.

Effect	E1: Desktop	E2: AR-HMD	E3: VR
Overall Cue Benefit	Very large RT effect (*d* = 1.19). Faster with cues.	Very large RT effect (*d* = 1.27). Faster with cues.	Large RT effect (*d* = 0.93). Faster with cues.
	Medium error effect (*d* = 0.67). Error decreases with cues.	Large error effect (*d* = 0.89). Error decreases with cues.	Large error effect (*d* = 0.99). Error decreases with cues.
Single *vs* Dual-Cue Benefit	Large RT effect (*d* = 0.89). Faster with dual cues.	Very large RT effect (*d* = 2.0). Faster with dual cues.	Very large RT effect (*d* = 1.11). Faster with dual cues.
	Medium error effect (*d* = 0.44). Lower error with dual cues.	Medium error effect (*d* = 0.62). Lower error with dual cues.	Medium error effect (*d* = 0.53). Lower error with dual cues.
Reliability Effect	Medium RT effect (*d* = 0.47). Faster with perfect cues.	Small - Medium RT effect (*d* = 0.31). Faster with perfect cues.	Medium RT effect (*d* = 0.61). Faster with perfect cues.
	Large error effect (*d* = 0.92). Lower error with perfect cues.	Very large error effect (*d* = 1.41). Lower error with perfect cues.	Large error effect (*d* = 0.90). Lower error with perfect cues.
Automation Bias	Large RT effect (*d* = 0.72). Faster with correct cues.	Very large RT effect (*d* = 1.20). Faster with correct cues.	Very large RT effect (*d* = 1.06). Faster with correct cues.
	Large error effect (*d* = 1.10). Lower error with correct cues.	Very large error effect (*d* = 1.42). Lower error with correct cues.	Very large error effect (*d* = 1.64). Lower error with correct cues.
Redundancy Gain	Negligible RT effect (*d* = 0.07)	Medium RT effect (*d* = 0.47). Faster with Global+Local cues.	Negligible RT effect (*d* = 0.05)
	No error effect (*d* = 0.19, p = .23)	No error effect (*d* = 0.26, p = .17)	No error effect (*d* = 0.01, *p* = .82)
Attentional Tunneling	No significant RT effect	No significant RT effect	No significant RT effect
	No significant error effect, but pattern of results showing higher error in HP when routine is cued	No significant error effect, but pattern of results showing higher error in HP when routine is cued	No significant error effect, but pattern of results showing higher error in HP when routine is cued
FOV Effect	NA	Faster when in of FOV (d = 0.56) and faster with G-L cues (d = 1.08)	Faster when in of FOV (d = 1.94) and faster with local cues (d = 1.02)
	NA	Lower error with G-L cue (d = 0.25).	Greater cost when cue is out of FOV which is amplified by global cue (d = 0.27)

Across all three experiments, H_1 was confirmed and replicated in prior work (Warden et al., 2022; Warden et al., 2023) showing an overall benefit of cueing compared to an unaided search (*ds* > 0.90). Cueing aids improved search efficiency and accuracy across all platforms. When directly comparing the overall cueing benefits across all three experiments, Experiment 3 (VR) had a significantly greater search time cueing benefit than Experiment 1 (Desktop; *d* = 0.80) and a similar search time cueing benefit as Experiment 2 (AR-HMD; *d* = 0.05). The largest search time cueing benefit was found using the AR-HMD (*d* = 1.27). For response time, cueing led to faster searches, especially when the search scene was constrained by the FOV or an HMD (Experiment 2). While all experiments show a medium to large effect of improved accuracy with cueing, the dynamic search task presented in VR resulted in the greatest benefit (*d* = 0.99), suggesting that cues facilitate accuracy the most in more realistic, complex, large, and dynamic search scenes where people must move forward to find objects.

Experiment 1 (desktop) confirmed that search time increased as the set size increased in both the cued and uncued conditions. That Experiment 2 (set size = 18) had a faster search time than Experiment 3 (set size = 22) also supports that set size plays a role in increased search time. These findings align with the self-terminating serial search (STSS) model, which predicts that search time increases linearly with set size (Neisser, 1964). The unaided search task (Figure 6) shows a more drastic increase in search time as the set size increases, indicative of a serial rather than a parallel search (Wickens et al., 2022). While this suggests a serial search in the case of the unaided condition, there is evidence that other factors play a role in search performance when we compare the unaided search to the cued searches.

First, the difference in search time across set sizes is less drastic in the cued condition (Figure 6). Second, all spatial cues led to faster and more accurate searches than the no-cue, and the icon cue in Experiment 2 led to a faster and more accurate search than the no-cue condition (see Tables 5 through 6). Consistent with Wolfe's guided search model (2021), the searcher's expectations of where to look (top-down processes) and the stimulus-driven information (bottom-up processes) provided by the cue also influence search performance by facilitating a more parallel search than a strict serial search. Spatial cues that directly indicate the target (i.e., arrow, gaze guidance line, highlight) provide salient information that exploits bottom-up processing effectively (Wolfe, 2021) unlike the minimap or icon cues that require some level of top-down processing. That the more stimulus driven cues led to the most effective searches have practical implications for design guidelines.

Additionally, and confirming H_2, the performance benefit of dual cueing was replicated across all three experiments. Dual cueing, specifically a global and local cue pair, consistently led to much faster response times and somewhat lower error rates compared to a single cue. The AR-HMD platform benefited the most from dual cueing for both response time (d = 2.0) and error (d = 0.62) compared to the 2D desktop (d_{RT} = 0.89; d_{ER} = 0.44) and VR (d_{RT} = 1.11; d_{ER} = 0.53) platforms. That the AR-HMD platform benefited the most from cueing, particularly dual cueing, is likely due to the constraints of the device's FOV, which prevents the searcher from using peripheral vision to assist in finding objects. Instead, head movements are required to bring information into the FOV, but without a cue or with only a single cue, the user does not initially know which direction to look for the target if that target is outside of the device's FOV.

Both the desktop and VR search experiments allow for the use of some peripheral vision during searches when deciding on the initial head turn direction, because their functional visual fields (i.e., foveal and peripheral regions) are not constrained by an actual HMD worn by the searcher, allowing for the attentional capture mechanism to vary across the experiments. In the desktop or VR experiment, the local cues (i.e., local arrow, highlight, gaze guidance line) can cue a target in the periphery and, hence, act as peripheral cues (see Jonides, 1981), which capture overt attention by automatically drawing the eyes and head towards the salient, bottom-up cue (Jonides, 1981). However, in the AR-HMD, this information is not available, and where to move the eyes or turn the head becomes a decision the searcher must make without such peripheral cue information. The fact that dual cues led to improved search performance suggests that they provide more guidance in the context of searches where peripheral information is unavailable (i.e., when the FOV is constrained). This aligns with the guided search model which suggests that more guidance results in more efficient searches (Wolfe, 2021), especially when the FOV is constrained, and the use of peripheral vision is limited.

It should be noted that the gaze guidance line cue was perpetually the better cue of all cues. This can be attributed to the cue possessing characteristics of both the global and local cues. The gaze guidance line was classified as a single cue because it was not paired with an additional cue. But it functioned like a global because where the cue pointed always remained in the searcher's field of view (i.e., the general direction was always indicated by the line that connected to the actual target). The gaze guidance line cue also functioned like a local cue because the line linked directly to the target object, signaling its precise location. This dual-like property of the single gaze guidance line cue likely contributed to its better performance compared to all other cues, especially the local cues.

While search performance was better with cues than without, the reliability of the automated cueing aid negatively impacted performance. As predicted by H_{1A}, accuracy degraded when the automation erred on occasion with the imperfect cueing aids. This effect was replicated across all three experiments, with the AR-HMD platform showing the largest cost to accuracy when the automation erred (see Table 7). The loss in accuracy is often associated with an automation bias where participants blindly follow the recommendation of the automation when the automation is wrong (Skitka et al., 1999; Warden et al., 2023). Both Experiments 1 and 3 show an overall cost to response time when automation erred. This increase in response time when automation was imperfect could suggest that people did not entirely blindly follow the advice of the automation or that there was some hesitancy in selecting the target objects (Warden et al., 2023). Experiment 2 (AR-HMD), on the other hand, failed to show a cost to response time, suggesting that, in that experiment exclusively, people were not taking additional time to check the real world data when the automation erred. This is further confirmed by the overall large loss in accuracy, suggesting a strong automation bias. Once again, it may be the constraints of the device's FOV that caused people to defer more to the automation rather than check the real world information.

Contradicting H_{2A}, there were no negative effects of imperfect reliability on dual-cues in either Experiment 1 or 2. Performance with imperfect dual-cues was hindered to the same degree as single-cues. Contrary to prior work where the more effective cue when automation is perfect, the worse it becomes when automation errs (Sebok & Wickens, 2017; Warden et al., 2023), the best cue in the present work did not suffer to a greater extent. This speaks favorably to dual-cueing systems, which do provide an overall cueing benefit but also do not induce a greater overall automation bias. However, in Experiment 3 the pattern of data, while not statistically significant, suggests that dual-cues did slightly amplified the automation bias. The similar pattern of results in each experiment may be attributed to the fact that there were so few trials for which the automation could error, resulting in low statistical power, an issue in the imperfect automation literature as a whole.

After pooling together the data from all three experiments (i.e., increasing statistical power), results did reveal that dual cueing showed an overall greater automation bias than single cues, and the magnitude of the effect was small ($d = 0.15$). That the best cue (dual cues) might show a greater automation bias in the VR experiment might be attributed to the more compelling and realistic nature of the task. Alternatively, it may be the case that the complexity of the VR scene, which simulated a real-world search task with dynamic movement, may have caused participants to offload the workload imposed by the search to the cueing aids more than the static and less realistic search tasks.

Hypothesis 3 predicted that cues providing both global and local spatial information would show a redundancy gain, such that performance would be better than either global or single cue alone. This was only confirmed once, for response time, in Experiment 2. Here, the redundant cue reduced search time more than either the global or local cues alone. However, no such redundancy gain was found for either response time or accuracy in Experiments 1 and 3. The local cue, in general, provided sufficient and equivalent information as the global and local combined cue to orient attention to the target. For these experiments, that performance measures were equivalent for the local and global+local cues and both being much better than the global cue alone, does beg the question as to whether the global cue is providing some kind of preattentive signal to orient a specific way until the effectiveness of the local cue takes over when searching for the target in the scene.

Hypotheses 1B and 2C predicted that cueing in general, and dual cueing in particular, would induce attentional tunneling such that performance on locating the less expected high priority target would degrade. Across the three experiments, there was weak evidence of attentional tunneling when the routine target was cued. Furthermore, when the data from all three experiments was combined, there was still weak evidence of attentional tunneling. But, for error data, the pattern of results was consistent and was indicative of the attentional tunneling phenomena found in previous literature (Yeh et al., 1999; Warden et al., 2023). While not statistically significant, errors were always higher for the less expected high priority target when the routine target was cued compared to not cued. Like with the imperfect automation trials, there were few instances when a high priority target could show up. Consequently, there is low statistical power and a large amount of variance in the data which might be masking the true attentional tunneling phenomena. When looking at the dual versus single cues, they both perform similarly across all three experiments which speaks favorably for the dual-cueing system. More specifically, while there is some evidence of attentional tunneling in general, the current work does not find evidence that dual cueing enhances attentional tunneling.

Hypothesis 4A and 4B were specific to only Experiments 2 and 3 where cues could be displayed in either world or screen-referenced coordinates. Confirming H_{4A}, single or dual cues that consisted of a world-reference or AR coordinate outperformed screen-referenced cues both in response time and accuracy. Furthermore, the dual-cues were better than the single, world-referenced cues, suggesting another benefit of dual cueing. Disconfirming H_{4B} and contradicting prior work (Warden et al., 2023), the more compelling cues (i.e., world-reference cues) failed to show a greater degree of cognitive tunneling compared to other cues.

Between Experiment Comparisons

There were notable commonalities and differences across the three platforms used in the current experiments. First, the search tasks differed in their complexity across each platform. The desktop static search task was the simplest form of search given the 2D images and the smallest FOV, which allowed for the exploitation of peripheral vision during searches. The AR-HMD search was more difficult due to the limited FOV of the device, which prevented the use of peripheral vision and, therefore, created a greater need for cues to know which direction to turn the head when a target fell outside of the immediate FOV of the device. Another reason the AR-HMD search task was more difficult can be attributed to integrating virtual cues presented on the HMD with the far domain search scene. Lastly, the dynamic search task in VR was the most difficult because the search required moving in space while integrating cue information on a simulated HMD with the realistic 3D scene beyond. Despite these differences in difficulty, this further speaks to when cueing, particularly dual cueing, may be more or less beneficial. For example, in the case of the AR-HMD where peripheral vision is constrained by the device, dual cues inform the user which way to turn their head to find the local cue that points to the target. This resulted in a redundancy gain for the global-local cue and improved search time more than either cue alone unlike the other experiments where peripheral vision was not constrained.

Across all three experiments, there was an overall cueing benefit compared to an unaided search. Furthermore, there was an overall benefit of dual-cueing. However, in Experiments 1 and 3 this benefit was driven by the local cue alone, unlike Experiment 2 where there was a redundancy gain for the global-local cue. Again, the desktop (Experiment 1) and VR (Experiment 3) search tasks did not constrain the searcher's FOV like the AR-HMD. When the user searches for the target on the desktop or in the VR environment, they could exploit their peripheral vision. This is partly because the target objects were never outside of the peripheral vision field of view to such a degree that they would have to make a head turn to notice the locally cued target. That the AR-HMD (Experiment 2) shows an amplified dual cueing benefit and redundancy gain for search time suggests that dual-cueing likely benefits scenarios where the target is not within the immediate FOV; that is, when the target is outside of the FOV, the global cue informs the searcher which direction to look to find the precise local cue. Without this information, the searcher must decide where they will look, which increases search time.

While the AR-HMD showed the greatest cueing benefit, the VR experiment showed a greater cueing and dual cueing benefit relative to the desktop display. The VR environment requires users to move through a large, 3D space to find the target, which increases the cognitive load of the search task. Given this added complexity, the search task benefits from cueing to a greater extent than the static 2D, smaller FOV search task presented on the desktop display. In the context of dual cueing, the benefits are also greater in VR than in the desktop search scene, which can, again, be attributed to the larger FOV in the VR environment. The dual cues can provide both general direction (global cue) and precise location (local cue), significantly improving search efficiency by reducing the need for scanning in the VR scene. This magnifies the cueing benefit because scanning in the smaller desktop display takes less time, especially when peripheral vision is exploited. Designers should implement dual-cueing in systems where targets appear outside of the FOV, where peripheral vision cannot be exploited. Dual cues in this situation will help direct the searcher's head movements to bring the locally cued target into view, thus compensating for any limitations of the FOV.

Another reason that cueing benefits differ across platforms can be attributed to the type of depth cues conveyed in the experiments. Experiment 1 had only 2D pictorial depth cues, the AR-HMD had a mix of some depth cues signaled by the HMD cues, and the VR experiment had 3D depth cues. Pictorial depth cues, like shading, perspective, and relative size, provide limited spatial information because they rely on flat, 2D images to imply depth information. This leads to an easier search because less cognitive effort is required to process limited-depth cue information. In contrast, 3D depth cues, like binocular disparity, represent more realistic spatial relationships between objects in the scene and the user but require more complex cognitive processing. Therefore, 3D depth cues in the scene require more mental effort and cognitive load when processing multiple layers of spatial information and navigating through an environment to identify targets. While the AR-HMD does not provide the same level of 3D complexity, there is still more spatial information than the pictorial cues alone. For example, the attention cues presented in the HMD provide a sense of depth between the display and the far domain search scene. Consequently, attention cues in more immersive environments with a greater degree of depth cues help mitigate increases in cognitive demand and therefore show an overall greater cueing benefit relative to an unaided search and relative to easier searches (i.e., desktop versus VR). Designers should consider the type of depth cues provided by the environment and how those cues interact with visual search.

The greatest automation bias was observed in the VR search task followed by the AR-HMD search task and lastly the desktop search task. The VR environment's higher level of complexity (i.e., 3D scene, dynamic movement) and interactivity may have led to greater reliance on the automation, resulting in stronger automation bias. The immersive nature of VR can make the attention cues feel more realistic and integrated and, perhaps, more trustworthy causing people to become over-reliant on the automation, even when the automation is incorrect. This aligns with prior work which has demonstrated that the most compelling and realistic cues led to a greater automation bias (Yeh et al., 1999; Warden et al., 2023). The realism of the cues in the current work increases from Experiment 1 to 2 to 3, which aligns with the observed increase in the magnitude of the automation bias. This finding highlights the importance of considering the interaction between the complexity of the search environment and the realism of the cues when designing an automated attention guidance system. Designers should consider exploring methods for mitigating automation bias, particularly for a more immersive and interactive environment, such as adding a level of transparency to the reliability of the cue or using more than one cue as a method for conveying certainty that the cued target is the correct target.

Limitations and Future Directions

While the current work provides valuable insights into cue effectiveness and dual cueing, it is important to address limitations of the work. One limitation is that the population studied consisted of undergraduate college students. Future work should seek to examine the implications of attention cueing with domain experts, such as soldiers or pilots. While the present work explores large search scenes, targets never appeared behind, above or below the searcher. Future work should explore the specific field of view threshold for which dual-cueing remains helpful by assessing the effects of dual cueing in even larger search scenes with wider search FOVs, such as when the target is behind, above, or below the searcher.

Future work should expand beyond the use of visual dual-cues by assessing when different types of dual cues (i.e., auditory-visual cues versus visual-visual cues) are more or less helpful for specific contexts. Additionally, that dual cueing shows a slight increase in a bias to follow the automation suggests that future research should explore methods for mitigating the automation bias, particularly for more immersive and interactive environments. Future work should explore the automation bias at different cue accuracies to develop a threshold that justifies using the automated system, and should investigate the role of more than two cues, particularly in conveying the amount of evidence that the cueing system has cued the correct target. Lastly, future work should assess under what circumstances cueing would not be helpful, and how kind of errors the system makes (i.e., false alarms versus misses) impacts its usefulness.

Conclusion

Conclusively, the current experiments demonstrate the overall effectiveness of duel-cueing systems in significantly enhancing search efficiency and accuracy in both static and dynamic visual search paradigms. The overall effectiveness of cueing was most notable when using the AR-HMD, which constrains the use of peripheral vision due to the limited field of view imposed by the device unlike the desktop and virtual environment search tasks. When the FOV is constrained, the searcher benefits from a global cue that signals which way to turn their head to find the local cue. Importantly, imperfect cueing aids hindered search accuracy indicative of an automation bias, and dual cueing slightly amplified this bias compared to single cues. Dual cueing also did not enhance attentional tunneling. Both of these findings suggest an advantage of dual cueing when the searcher's FOV is constrained and peripheral vision limited. Overall, these findings highlight the critical role of cue properties in the design of attentional cueing systems used for static 2D displays, and static and dynamic searches conducted with an HMD. These findings have design guideline implications for cueing systems that are imperfectly reliable.

REFERENCES

Azuma, R. (1997). A survey of augmented reality. Teleoperators and Virtual Environments, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. IEEE Computer Graphics and Applications, *21*(6), 34–47. 10.1109/38.963459

Baird, K. M., & Barfield, W. (1999). Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual reality*, *4*, 250-259. https://doi.org/10.1007/BF01421808

Bauerfeind, K., Drüke, J., Schneider, J., Haar, A., Bendewald, L., & Baumann, M. (2021). Navigating with Augmented Reality–How does it affect drivers' mental load?. *Applied Ergonomics*, *94*, 103398. https://doi.org/10.1016/j.apergo.2021.103398

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of cognitive neuroscience*, *7*(1), 66-80. https://doi.org/10.1162/jocn.1995.7.1.66

Billinghurst, M., Bowskill, J., Dyer, N., & Morphett, J. (1998, March). An evaluation of wearable information spaces. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium, (Cat. No.98CB36180)*, Atlanta, GA, USA, 20-27. doi: 10.1109/VRAIS.1998.658418.

Biocca, F., Owen, C., Tang, A., & Bohil, C. (2007). Attention issues in spatial information systems: directing mobile users visual attention using AR. Journal of management information systems, *23*(4), 163-184. https://doi.org/10.2753/MIS0742-1222230408

Binetti, N., Wu, L., Chen, S., Kruijff, E., Julier, S., & Brumby, D. P. (2021). Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. *Displays*, *69*, 102032. https://doi.org/10.1016/j.displa.2021.102032

Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557. https://doi.org/10.7334/psicotherma2016.3783

Boston, B. N., & Braun, C. C. (1996). Clutter and display conformality: Changes in cognitive capture. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(2), 57-61. https://doi.org/10.1177/154193129604000211

Caird, J. K., Horrey, W. J., & Edwards, C. J. (2001). Effects of conformal and nonconformal vision enhancement systems on older-driver performance. *Transportation Research Record*, *1759*(1), 38-45. https://doi.org/10.3141/1759-05

de Oliveira Faria, N., Gabbard, J. L., & Smith, M. (2020). Place in the world or place on the screen? Investigating the effects of augmented reality head-up display user interfaces on drivers' spatial knowledge acquisition and glance behavior. In *2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, 762-763. 10.1109/VRW50115.2020.00232

Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. E. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Frontiers in Robotics and AI, *5*, 37. https://doi.org/10.3389/frobt.2018.00037

Ellis, S. R. (2006). *Towards determination of visual requirements for augmented reality displays and virtual environments for the airport tower* (No. AD-A473306). NATO R & T Organization

Explosive hazards. (2024). CAT-UXO. https://cat-uxo.com/explosive-hazards

Fadden, S., Wickens, C. D., & Ververs, P. (2000). Costs and benefits of head up displays: An attention perspective and a meta analysis. *SAE Technical Papers Series*. https://doi.org/10.4271/2000-01-5542

Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. In *Proceedings of the human factors and ergonomics society annual meeting*, *42*(1), 16-20. https://doi.org/10.1177/154193129804200105

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: are they viable?. *Human Factors*, *43*(2), 173-193. https://doi.org/10.1518/001872001775900841

Galster, S. M., Bolia, R. S., Roe, M. M., & Parasuraman, R. (2001). Effects of automated cueing on decision implementation in a visual search task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 321-325.

Gruenefeld, U., Löcken, A., Brueck, Y., Boll, S., & Heuten, W. (2018). Where to look: Exploring peripheral cues for shifting attention to spatially distributed out-of-view objects. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 45(4), 221-228. https://doi.org/10.1177/154193120104500412

Greenwood, P. M., & Parasuraman, R. (2004). The scaling of spatial attention in visual search and its modification in healthy aging. *Perception & psychophysics*, *66*, 3-22. https://doi.org/10.3758/BF03194857

Greenwood, P. M., & Parasuraman, R. (1999). Scale of attentional focus in visual search. *Perception & Psychophysics*, *61*(5), 837-859. https://doi.org/10.3758/BF03206901

Greenwood, P. M., Parasuraman, R., & Alexander, G. E. (1997). Controlling the focus of spatial attention during visual search: effects of advanced aging and Alzheimer disease. *Neuropsychology*, *11*(1), 3. https://doi.org/10.1037/0894-4105.11.1.3

Heidel, R. E. (2016). Causality in statistical power: Isomorphic properties of measurement, research design, effect size, and sample size. *Scientifica (Cairo)*, 8920418. https://doi.org/10.1155/2016/8920418

Hein, P., Bullinger, A. C., & Bernhagen, M. (2020). Two is better than one: Improved attention guiding in AR by combining techniques. IEEE Computer Society, *40*(5), 57-66. doi: 10.1109/MCG.2020.3012274

Helleberg, J. R., & Wickens, C. D. (2003). Effects of Data-Link Modality and Display Redundancy on Pilot Performance: An Attentional Perspective. *The International Journal of Aviation Psychology*, *13*(3), 189–210. https://doi.org/10.1207/S15327108IJAP1303_01

Henderson, S., & Feiner, S. (2010). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE transactions on visualization and computer graphics*, *17*(10), 1355-1368. doi: 10.1109/TVCG.2010.245

Jeffri, N. F. S., & Rambli, D. R. A. (2021). A review of augmented reality systems and their effects on mental workload and task performance. *Heliyon*, *7*(3). https://doi.org/10.1016/j.heliyon.2021.e06277

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movements. In J. Long & A. Baddeley (Eds.), Attention and performance IX (pp. 187–203). Lawrence Erlbaum Associates, Inc.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological review*, *96*(1), 101-124.

Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). *An Operational evaluation of head up displays for civil transport operations. NASA/FAA phase 3 report* (No. NASA-TP-1815).

Lebeck, K., Ruth, K., Kohno, T., & Roesner, F. (2017). Securing augmented reality output. In *2017 IEEE symposium on security and privacy (SP)*, 320-337. doi: 10.1109/SP.2017.13

Kim, K. H., & Wohn, K. Y. (2011). Effects on productivity and safety of map and augmented reality navigation paradigms. *IEICE Transactions on Information and Systems*, *94*(5), 1051-1061. 10.1587/transinf.E94.D.1051

Kim, H., Miranda Anon, A., Misu, T., Li, N., Tawari, A., & Fujimura, K. (2016). Look at me: Augmented reality pedestrian warning system using an in-vehicle volumetric head up display. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 294-298. https://doi.org/10.1145/2856767.28568

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver behavior and performance with augmented reality pedestrian collision warning: An outdoor user study. *IEEE transactions on visualization and computer graphics*, *24*(4), 1515-1524. doi: 10.1109/TVCG.2018.2793680

Kumaran, Milner, Bullock, Giesbrecht, and Hollerer. (2023). The Impact of Navigation Aids on Search Performance and Object Recall in Wide-Area Augmented Reality. CHI, 710, 1-17. doi.org/10.1145/3544548.3581413

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 57-87. https://doi.org/10.1177/1555343411433844

Martin-Emerson, R., & Wickens, C. D. (1992). The vertical visual field and implications for the head-up display. *Proceedings of the Human Factors Society Annual Meeting*, *36*(18), 1408–1412. https://doi.org/10.1177/154193129203601810

Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Nasanen, J., & Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. *In Proceedings of the SIGCHI conference on human factors in computing systems*, 1889-1898. https://doi.org/10.1145/1518701.1518991

Naish, J. M. (1964). Combination of information in superimposed visual fields. *Nature*, *202*(4933), 641-646. https://doi.org/10.1038/202641a0

Neisser, U. (1964). Visual Search, Scientific American, 210, 94-102. https://www.jstor.org/stable/24931530

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. In *Decision Making in Aviation*, 251-260. Routledge.

Peereboom, J., Tabone, W., Dodou, D., & de Winter, J. (2023). Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios. *Virtual Reality, 28*, 119. https://doi.org/10.1007/s10055-024-01017-9

Poole, C. A., Warden, A. C., Wickens, C. D., Raikwar, A., Clegg, B. A., & Ortega, F. R. (2023). Information access costs with a wide-angle desktop display. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 947-952. https://doi.org/10.1177/21695067231192646

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Raikwar, A., Mifsud, D., Wickens, C. D., Kelly, B., Warden, A. C., Clegg, B. A., & Ortega, F. R. (2023, October). Beyond the Wizard of Oz: Using Imperfect Machine Learning to Examine the Impact of Reliability of Augmented Reality Cues on Visual Search Performance. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 437-438. doi: 10.1109/ISMAR-Adjunct60411.2023.00093

Renner, P., & Pfeiffer, T. (2017, March). Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE symposium on 3D user interfaces (3DUI)*, 186-194. 10.1109/3DUI.2017.7893338

Robertson, C. M., MacIntyre, B., & Walker, B. N. (2008). An evaluation of graphical context when the graphics are outside of the task area. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 73-76. doi: 10.1109/ISMAR.2008.4637328

Rusch, M. L., Schall Jr, M. C., Gavin, P., Lee, J. D., Dawson, J. D., Vecera, S., & Rizzo, M. (2013). Directing driver attention with augmented reality cues. *Transportation research part F: traffic psychology and behaviour*, *16*, 127-137. https://doi.org/10.1016/j.trf.2012.08.00

Schankin, A., Reichert, D., Berning, M., & Beigl, M. (2017). [POSTER] The Impact of the Frame of Reference on Attention Shifts Between Augmented Reality and Real-World Environment. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, 25-30. doi: 10.1109/ISMAR-Adjunct.2017.24

Schinke, T., Henze, N., and Boll, S. (2010). Visualization of off-screen objects in mobile augmented reality. *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services,* 313-316. https://doi.org/10.1145/1851600.1851655

Schmerwitz, S., Lueken, T., Doehler, H. U., Peinecke, N., Ernst, J. M., & da Silva Rosa, D. L. (2017). Conformal displays: human factor analysis of innovative landing aids. *Optical Engineering*, *56*(5), 051407-051407. https://doi.org/10.1117/1.OE.56.5.051407

Sebok, A., & Wickens, C. D. (2017). Implementing lumberjacks and black swans into model-based tools to support human–automation interaction. *Human factors*, *59*(2), 189-203. https://doi.org/10.1177/0018720816665201

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making?. *International Journal of Human-Computer Studies*, *51*(5), 991-1006. https://doi.org/10.1006/ijhc.1999.0252

Soret, Hurter, Charras, Peysakhovich. (2019). Attentional orienting in VR using endogenous and exogenous cues in auditory and visual modalities. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, 86*, 1-8. https://doi.org/10.1145/3317959.3321490

Sternberg, S. (1966). High speed scanning in human memory, Science, *153*(3736), 652-654. DOI: 10.1126/science.153.3736.652

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, *12*(6), 462-466. https://doi.org/10.1111/1467-9280.00386

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 73-80. https://doi.org/10.1145/642611.642626

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 12. https://doi.org/10.1037/0096-1523.10.1.12

Ververs, P. M., & Wickens, C. D. (2000). Designing head-up displays (HUDs) to support flight path guidance while minimizing effects of cognitive tunneling. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, *44*(13), 45-48. https://doi.org/10.1177/154193120004401312

Warden, A. C., Wickens, C. D., Clegg, B. A., Rehberg, D., & Ortega, F. R. (2022a). Information access effort: The role of head movements for information presented at increasing eccentricity on flat panel and head-mounted displays. *Human Factors*, *66*(8), 2057-2081. https://doi.org/10.1177/154193120004401312

Warden, A. C., Wickens, C. D., Mifsud, D., Ourada, S., Clegg, B. A., & Ortega, F. R. (2022b). Visual Search in Augmented Reality: Effect of Target Cue Type and Location. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66*(1), 373-377. https://doi.org/10.1177/107118132266

Warden, A. C., Wickens, C. D., Rehberg, D., Ortega, F. R., & Clegg, B. A. (2023). Fast, Accurate, but Sometimes Too-Compelling Support: The Impact of Imperfectly Automated Cues in an Augmented Reality Head-Mounted Display on Visual Search Performance, *IEEE Transactions on Human-Machine Systems, 53*(6), 1061-1072. doi: 10.1109/THMS.2023.3302152

Warden, A. C., Wickens, C. D., Clegg, B. A., & Ortega, F. R. (2024). Quantitative effects of overlay clutter and information access effort: Examining the scan-clutter trade-off in displays with geospatial maps. *Journal of Experimental Psychology: Applied.* Advance online publication. https://doi.org/10.1037/xap0000512

Wickens, C. D., & Long, J. (1994). Conformal symbology, attention shifts, and the head-up display. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(1), 6-10. https://doi.org/10.1177/154193129403800103

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. (2005). Attentional tunneling and task management. International Symposium on Aviation Psychology, 812–817.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201-212. https://doi.org/10.1080/14639220500370105

Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. *Human Factors*, *51*(5), 638-651. https://doi.org/10.1177/0018720809349709

Wickens, C. D., McCarley, J. S., & Gutzwiller, R. S. (2022). *Applied attention theory*. CRC press.

Wither, J., DiVerdi, S., & Hollerer, T. (2007). Evaluating display types for ar selection and annotation. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 95-98. doi: 10.1109/ISMAR.2007.4538832

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136*(4), 623–638. https://doi.org/10.1037/0096-3445.136.4.623

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060-1092. https://doi.org/10.3758/s13423-020-01859-9

Wu, C. C., & Wolfe, J. M. (2019). Eye movements in medical image perception: a selective review of past, present and future. *Vision*, *3*(2), 32. https://doi.org/10.3390/vision3020032

Yeh, M., Wickens, C. D., & Seagull, F. J. (1998). Conformality and target cueing: Presentation of symbology in augmented reality. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(21), 1526-1530. https://doi.org/10.1177/154193129804202113

Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human factors*, *45*(3), 390-407. https://doi.org/10.1518/hfes.45.3.390.27249

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*(4), 524-542. https://doi.org/10.1518/001872099779656752

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, *43*(3), 355-365. https://doi.org/10.1518/001872001775898269

Zhao, Y., Stefanucci, J., Creem-Regehr, S., & Bodenheimer, B. (2023). Evaluating Augmented Reality Landmark Cues and Frame of Reference Displays with Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, *29*(5), 2710-2720. doi: 10.1109/TVCG.2023.3247078

APPENDIX A: EXPERIMENT 1 TASK INSTRUCTIONS

Thank you for agreeing to participate in this experiment! Before you begin, you will read the information below about the experiment and the task.

Please carefully read through the instructions below:

You will complete a visual search task using a wide-angle desktop computer. During the task, you will locate target objects embedded in real-world images. Target objects consist of different kinds of rocks, plastic bottles, metal objects, and logs. Before each trial, you will have 5 seconds to carefully study the target object that you need to locate since its image will not always be available to you after the search starts.

You will complete the search with and without cues to assist in the task. Trials with a cue will cue you to the location of the regular target object studied before each trial. In addition to this target object, there will also be a high priority object on some of the trials. High priority objects consist of different kinds of improvised explosive devices (IED)s. High priority objects are **never cued** and should take precedence over regular targets.

Once you find the regular target and also the high priority target if it may be present, you will use the mouse left click to select the target(s) for each trial. When you select the target(s) they will disappear. You can only make a maximum of 2 selections per trial. Once you think you are done making your selection, you will press the CTRL (control) button on the keyboard.

On cued trials, you will use different types of cues displayed on the monitor. You will use either one or two cues to assist in the search task.

- The **Icon Cue** provides you an image of the target that you must find. This image will be displayed at the center of the screen until you locate the target.
- The **Minimap Cue** will show you a map of where all the objects in the room are located. The target that you are searching for will highlighted with a yellow outline.
- The **Global Arrow Cue** points in the direction of the target that you are searching.
- The **Local Arrow Cue** is an arrow that is positioned directly above the target object and points down to it.

- The **Gaze Guidance Cue** is a yellow line that projects from the center of the screen to the target.
- The **Highlight Cue** is a yellow highlight that appears around the target in the search scene.

For trials with two cues, you will use a combination of the two cues above. For example, the **Minimap Cue** and the **Local Arrow Cue** provide information about the targets location. For the **No Cue** condition you will be required to find the object without an aid to assist you.

The Icon, Minimap, and Global Arrow cues will be located in the center of the display. You will complete multiple trials for each cue. Before you start, you will be informed which cue you will be getting.

Prior to the main trials, you will complete practice trials to familiarize you with each cue condition and how to make your selection. If you have ANY questions after the practice trials, please let the experimenter know.

You will complete trials where the cueing aids may NOT BE PERFECTLY RELIABLE, meaning you may not fully rely on them. Before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

You will also complete trials where the cue is always correct. However, before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

As a reminder, there will be UNCUED high priority objects that appear periodically and should take precedence over regular targets. You will use the mouse (left click) to select the target(s). Once you have found all of the targets, you will push the Left CTRL (control) button on the keyboard. If you do not see the target or targets in the search field, you can also press the Left CTRL (control) button on the keyboard to indicate "object not present." Make all responses as rapidly and accurately as possible.

LET THE EXPERIMENER KNOW YOU ARE DONE READING THE TEXT AND READY TO SEE EXAMPLE OBJECTS.

Regular Target Objects

- You will be shown a regular target to search for at the beginning of each trial. You will see this target for 5 seconds at the beginning of each trial. Below are examples of the four types of regular targets you will have to search for:

Rocks

Plastic Bottles

Metal Items

Wood Items

High Priority Target Objects

- Periodically, there will also be High Priority targets that you should find in addition to the regular targets. High Priority targets take precedence. Below are what these objects look like:

APPENDIX B: EXPERIMENT 2 TASK INSTRUCTIONS

Thank you for agreeing to participate in this experiment! Before you begin, you will read the information below about the experiment and the task.

Please carefully read through the instructions below:

You will complete a visual search task using an Augmented Reality Head-Mounted Display. During the task, you will locate target objects embedded in real-world images. Target objects consist of different kinds of rocks, plastic bottles, metal objects, and logs. Before each trial, you will have 3 seconds to carefully study the target object that you need to locate since its image will not always be available to you after the search starts.

You will complete the search with and without cues to assist in the task. Trials with a cue will cue you to the location of the regular target object studied before each trial. In addition to this target object, there will also be a high priority object on some of the trials. High priority objects consist of different kinds of improvised explosive devices (IED)s. High priority objects are **never cued** and should take precedence over regular targets.

Once you find the regular target and also the high priority target if it may be present, you will use the **RIGHT SHIFT button** to select the target(s) for each trial. When you select the target(s) they will disappear. You can only make a maximum of 2 selections per trial. Once you think you are done making your selection, you will press the **LEFT CTRL (control)** button on the keyboard.

On cued trials, you will use different types of cues displayed on the device. You will use either one or two cues to assist in the search task.

- The **Icon Cue** provides you an image of the target that you must find. This image will be displayed at the center of the screen until you locate the target.
- The **Minimap Cue** will show you a map of where all the objects in the room are located. The target that you are searching for will highlighted with a yellow outline.
- The **Global Arrow Cue** points in the direction of the target that you are searching.
- The **Local Arrow Cue** is an arrow that is positioned directly above the target object and points down to it.

- The **Gaze Guidance Cue** is a yellow line that projects from the center of the screen to the target.
- The **Highlight Cue** is a yellow highlight that appears around the target in the search scene.

For trials with two cues, you will use a combination of the two cues above. For example, the **Minimap Cue** and the **Local Arrow Cue** provide information about the targets location. For the **No Cue** condition you will be required to find the object without an aid to assist you.

The Icon, Minimap, and Global Arrow cues will be located in the center of the display. You will complete multiple trials for each cue. Before you start, you will be informed which cue you will be getting.

Prior to the main trials, you will complete practice trials to familiarize you with each cue condition and how to make your selection. If you have ANY questions after the practice trials, please let the experimenter know.

You will complete trials where the cueing aids may NOT BE PERFECTLY RELIABLE, meaning you may not fully rely on them. Before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

You will also complete trials where the cue is always correct. However, before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

As a reminder, there will be UNCUED high priority objects that appear periodically and should take precedence over regular targets. You will use the RIGHT SHIFT key to select the target(s). Once you have found all of the targets, you will push the LEFT CTRL (control) button on the keyboard. If you do not see the target or targets in the search field, you can also press the Left CTRL (control) button on the keyboard to indicate “object not present.” Make all responses as <u>RAPIDLY and ACCURATELY</u> as possible.

LET THE EXPERIMENER KNOW YOU ARE DONE READING THE TEXT AND READY TO SEE EXAMPLE OBJECTS.

Regular Target Objects

- You will be shown a regular target to search for at the beginning of each trial. You will see this target for 5 seconds at the beginning of each trial. Below are examples of the four types of regular targets you will have to search for:

Rocks

Plastic Bottles

Metal Items

Wood Items

High Priority Target Objects

- Periodically, there will also be High Priority targets that you should find in addition to the regular targets. High Priority targets take precedence. Below are what these objects look like:

APPENDIX C: EXPERIMENT 3 TASK INSTRUCTIONS

Thank you for agreeing to participate in this experiment! Before you begin, you will read the information below about the experiment and the task.

Please carefully read through the instructions below:

You will complete a visual search task in Virtual Reality. During the task, you will simulate walking **forward** along a path to locate target objects embedded in the virtual scene. **<u>You will only be able to move in the forward direction from your starting point.</u>**

The objects will appear on either side of the trail and will consist of different kinds of rocks, bottles, cans, and logs. Before each trial, you will have 3 seconds to carefully study the target object that you need to find for that trial. Note that the image will not be available to you after the search starts.

You will complete the walking search task with and without cues to assist in the task. Trials with a cue will cue you to the location of the regular target object that you studied. In addition to this target object, there will also be a **high priority object** on some of the trials. High priority objects consist of different kinds of improvised explosive devices (IED)s. High priority objects are **never cued** and should take precedence over regular targets.

Once you find the regular target and also the high priority target if it may be present, you can line up the cross hair on the object until a box appears around it. This means that you can select that object. To select the target, you will use the **A Button on the Right Hand controller** to select the target(s) for each trial. When you select the target(s) they will disappear. You can only make a maximum of 2 selections per trial. Once you think you are done making your selection, you will press the **X Button on the Left Hand controller** to go to the next trial.

On cued trials, you will use different types of cues displayed on the device. You will use either one or two cues to assist in the search task.

- The **Minimap Cue** will show you a map of where all the objects in the room are located. The target that you are searching for will highlighted with a yellow outline.
- The **Global Arrow Cue** points in the direction of the target that you are searching.

- The **Local Arrow Cue** is an arrow that is positioned directly above the target object and points down to it.
- The **Gaze Guidance Cue** is a yellow line that projects from the center of the screen to the target.

For trials with two cues, you will use a combination of two cues above. For example, the **Minimap Cue** and the **Local Arrow Cue** provide information about the targets location. For the **No Cue** condition you will be required to find the object without a cue to assist you.

Prior to the main trials, you will complete practice trials to familiarize you with each cue condition and how to make your selection. If you have ANY questions after the practice trials, please let the experimenter know.

You will complete trials where the cueing aids may NOT BE PERFECTLY RELIABLE, meaning you may not fully rely on them. Before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

You will also complete trials where the cue is always correct. However, before selecting the object that is identified by the cue, you should try to ensure that the selected object is, in fact, the same as the one that you saw at the beginning of the trial.

As a reminder, there will be UNCUED high priority objects that appear periodically and should take precedence over regular targets.

You will use the <u>A Button on the Right Hand</u> controller to select the target(s). Once you have found all of the targets, you will push the <u>X Button on the Left Hand</u> controller to go to the next trial. If you do not see the target or targets in the search field, you can also press <u>X Button on the Left Hand</u> controller to indicate "object not present." Make all responses as <u>RAPIDLY and ACCURATELY</u> as possible.

LET THE EXPERIMENER KNOW YOU ARE DONE READING THE TEXT AND READY TO SEE EXAMPLE OBJECTS.

Regular Target Objects

- You will be shown a regular target to search for at the beginning of each trial. You will see this target for 3 seconds at the beginning of each trial. Below are examples of the four types of regular targets you will have to search for:

Rocks

Bottles

Metal Items

Wood Items

High Priority Target Objects

- Periodically, there will also be High Priority targets that you should find in addition to the regular targets. High Priority targets take precedence. Below are what these objects look like:

APPENDIX D: TRUST IN AUTOMATION QUESTIONNAIRE

In the following questions, "the system" refers to the automation. You will answer each question with a slider to indicate how much you trust the system on a scale ranging from 0 to 100%. Answer honestly based on your experience with this automation.

1. To what extent can the system's behavior be predicted from moment to moment?
2. To what extent can you count on the system to do its job?
3. What degree of faith do you have that the system will be able to cope with all system states in the future?
4. Overall, how high was your self-confidence in the automation?
5. Overall, how much do you trust the automation?
6. How much do you trust in the automation to find the correct object?
7. How much self-confidence do you have in your ability to find the correct object?

Now please rate your trust and self-confidence in each of the following:

1. Trust in the local bus service to get you to the store on time.
2. Self-confidence in your ability to get to the store on time.
3. Trust in your calculator or computer to produce the right answer.
4. Your self-confidence in your ability to arrive at the correct answer doing the calculations manually.
5. Trust in the heating system where you live to keep you comfortable.
6. Your self-confidence in your ability to turn the heater on and off manually to keep you comfortable.
7. Trust in your watch to tell the correct time.
8. Your self-confidence in your ability to estimate the correct time.

APPENDIX E: CUE USABILITY QUESTIONS

1. Did the cues help your performance?
2. What cue helped you find objects the most?
3. What cue helped the least when finding objects?
4. Did the single cues or the dual cues help your performance most?
5. Was the no cue condition hard? (range: easy, moderate, difficult)
6. What cue did you prefer most?
7. Have you used virtual reality before this experiment? (Only Experiment 3)