Flinders University

Cueing Induced Inattentional Blindness

By

David P. Nicoll
Bachelor of Psychological Science (Honours)

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy
College of Education, Psychology & Social Work

18th March 2024

Table of Contents

List of Tables

List of Figures

Summary

Augmented Reality (AR) technology is increasingly used as a visual search aid in operational search environments. In such contexts, visual search rarely involves a single well-defined target. Instead, target descriptors are necessarily vague, such as "search for any weapon", leading to both human and AR technology finding it challenging to identify all targets reliably. When AR technology cannot identify all targets reliably, the search task becomes a dual-search task. The first is for targets identified by the AR system, and the second is for targets not identified. In complex search environments, it would seem that providing any support to operators, such as cueing some targets, should aid their performance. However, prior research has demonstrated that performance improves for the cued search task, but performance on other concurrent tasks can suffer. In summary, making one task easier does not guarantee that other tasks will benefit.

This project adapted classic sustained inattentional blindness paradigms into a video game simulating a military overwatch scenario. During three experiments, participants performed a demanding primary task, where they were directed to click on any threats they saw. Depending on the experiment, trial and condition, people were aided by either an informative, uninformative, or no cue. Behavioural responses to cued and uncued threats were measured, with inattentional blindness measured as the miss-rate of uncued rare threats, which were never cued. Additionally, overt attention was recorded using eye-tracking for all experiments and a combination of both overt and covert attention was recorded using Steady-state Visually Evoked Potentials, using electroencephalography, for the last two experiments.

Experiment 1 was a within-subjects design with repeated onsets of the uncued rare threat character. Each participant experienced equal numbers of cued and uncued trials. When cues were provided, they were 100% informative in cueing frequent threat characters. Behavioural and eye-gaze measures were recorded. As expected, performance towards cued

threats improved, but without the predicted cost to uncued threat search performance. However, there was a general trend toward reduced uncued threat search performance, and sizable differences in attention towards cued characters.

Experiment 2 moved to a between-subjects design with only a single rare threat character onset. Additionally, an uninformative cue condition was introduced. Participants were assigned to only one of the conditions experiencing uncued, informatively cued, or uninformatively cued frequent threats only. Both cued conditions showed increased inattentional blindness rates compared to the uncued condition but with differing attentional patterns.

Experiment 3 replicated Experiment 2 as closely as practical but using a virtual reality system. No changes were made to the experimental paradigm. Informative cue results closely replicated that seen in Experiments 1 and 2. However, unlike the Experiment 2 results the uninformative cue condition outcomes were more similar to the no-cue condition results.

These findings support prior research indicating that providing task-relevant visual cues leads to tunnelling of attention towards those cues, impairing the search for uncued targets. In contrast, task-irrelevant cues contribute to visual clutter and, depending on the search context, may increase inattentional blindness.

Declaration

I certify that this thesis:

1. Does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university
2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and
3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed David Nicoll

Date 18/03/2024

List of Conference Abstracts and Presentations

Research from this thesis has been presented at:

David Nicoll, Salvatore Russo, Megan Bartlett, John Salamon, Mike Nicholls, Tobias Loetscher & Oren Griffiths (2023, April). *Attention distribution during a visual search task in screen-based versus virtual reality displays* [Poster presentation]. Experimental Psychology Conference, Canberra, Australia.

David Nicoll, Salvatore Russo, Megan Bartlett, John Salamon, Mike Nicholls, Tobias Loetscher & Oren Griffiths (2022, November). *Distribution of attention in screen-based versus head-mounted virtual reality displays in a military-relevant search task* [Poster presentation]. Defence Human Sciences Symposium, Paramatta, Australia.

David Nicoll, Sal Russo, Megan Bartlett, Mike Nicholls, Tobias Loetscher, Oren Griffiths (2022, July). *Impact of Augmented Reality visual cues on attention and search performance* [Conference presentation]. Australasian Brain & Psychological Sciences Meeting, Brisbane, Australia.

Salvatore Russo, David Nicoll, Megan Bartlett, Mike Nicholls and Oren Griffiths (2021, November). *Investigating the benefits and costs of target detection aids* [Conference presentation]. Defence Human Sciences Symposium, Melbourne, Australia.

David Nicoll & Oren Griffiths (2020, October). *Cueing induced Inattentional Blindness* [Colloquium presentation]. Psychology Seminar Series, Flinders University, Adelaide, Australia.

Acknowledgements

Dr Oren Griffiths. Thank you for your support, commitment, and time throughout my PhD. The variety of research experiences and skills that I have acquired, thanks to your commitment to my development, is astounding. Your persistence and honesty as a supervisor were always appreciated.

Professor Mike Nicholls. Being part of your Brain and Cognition Lab has been an honour and pleasure. At critical moments, you stepped in with support, knowledge and advice that was invaluable and very much appreciated. Thank you for your commitment to my success.

Dr Sal Russo. It would be remiss not to acknowledge the volume of difficulties we surmounted. The shared experiences of the COVID years, learning new programming languages, technology, and trying to get brains to oscillate will not be forgotten. I thoroughly enjoyed our time working together and will always appreciate the support and friendship you offered.

Drs Megan Bartlett and John Salamon. I am grateful to have worked with you both. Your friendship, knowledge and advice have been critical to my PhD success and the enjoyment of my time in the lab.

To my fellow PhD cohort, it has been a fantastic journey. Alycia, Steph and Tess, sharing the lab and your friendship will always be a treasured memory. Georgina, I strongly suspect that without you organising those study groups in undergrad, I would not have made it through to a PhD. Your friendship and eternal presumptions that I will do well were a great support. All of you have impressed and inspired me with how you have mastered the challenges of your PhD whilst dealing with life’s difficulties. I look forward to seeing where you will succeed next.

To my wife, Lynne, this undergrad and PhD journey would have been impossible without your support and belief in me. You are an impressive woman and an amazing partner. Your efforts and sacrifices over the last five years were significant, and I can only hope you truly understand my love and appreciation for you. To my family, thank you for your support and patience. It is humbling to think that three beautiful children were added to our family in the time I took to do this very big assignment. I am proud of you all.

I would also like to state my gratitude for the financial support I have received. This PhD was funded by Human Performance Research contract HPRnet ID9101 “Helped, not hindered: Optimising the use of augmented reality by Defence personnel”, awarded to the supervisors of this PhD; Professor Mike Nicholls and Dr. Oren Griffiths. Scholarships and awards from Flinders University and the College of Education, Psychology, and Social Work supported presenting my research at conferences.

Chapter 1: Literature Review

Augmented Reality and the Dismounted Soldier

The modern battlefield is an increasingly data-rich environment (Skinner, 2019). A dismounted soldier receives information from a network of sources: drones, surveillance aircraft, vehicles, weapons systems and other soldiers in the field. The volume and complexity of data streams soldiers receive in the field are only set to increase, and this is not a problem unique to the military. To minimise the invasiveness of surgical procedures, surgeons increasingly rely on techniques that constrict their direct view (e.g., keyhole surgery), increasing reliance on technology to guide and support the procedure (Dixon et al., 2013). In the military context, information is typically provided piecemeal on separate devices, potentially leading to increased response times and mental load as the soldier switches focus between these different sources of task-relevant information.

To address this problem, Augmented Reality (AR) technology is being investigated that presents task-relevant information directly into the dismounted soldier's field of vision, reducing the number of devices a soldier must attend to. However, as with all technological developments that seek to optimise one performance aspect, costs may be associated with the performance of other non-optimised functions. Consequently, it is critical to holistically examine the impact of the new technology on overall performance. For example, early body armour technology provided high levels of personal protection but at the cost of agility, reducing soldiers' ability to extricate themselves from potentially lethal situations (Fish & Scharre, 2018). Similarly, AR studies in various domains have found comparable cost/benefit trade-offs. In military aerial search, cueing the pilot to potential targets increased detection speed for most cued targets but resulted in blindness to vital uncued targets and even some cued targets (Yeh & Wickens, 2001). In a surgical context, Dixon et al. (2013) found that placing AR visual overlays of anatomical features on the surgeon's field of vision improved the accuracy of an endoscopy procedure but impaired their ability to detect potentially fatal surgical complications. These findings suggest that introducing technology such as an AR system that highlights friend vs foe may improve the performance of identifying highlighted individuals on the battlefield. Still, there could plausibly be some cost to the soldier's overall performance.

This project investigates the factors influencing the costs and benefits associated with AR cues designed to direct attentional deployment during operational visual search. To better understand any potential trade-off, it is important to consider the factors that affect the difficulty of visual search and the mechanisms by which visual cues drive attentional allocation. Thus, I initially considered cognitive theories of visual search and cueing.

Visual Search

Visual search is the process of looking for a particular target stimulus that is present amongst a set of non-target stimuli, often referred to as distractors. Searching for a pen (target) amongst all the other objects (distractors) on a desk is an example of a visual search. Visual search theories state that search is easier when a basic feature of the target object makes it stand out from the surrounding visual field, allowing for a rapid form of search called "parallel search" (for reviews, see Theeuwes, 2013; Wolfe, 1994). This search type is also commonly referred to as "pop-out" because the high salience of a stimulus pre-consciously and rapidly draws attention to itself from any location in the visual field, irrespective of the number of distractors present (Treisman, 1985). Examples include a single red circle amongst some blue circles or an object moving while other objects are stationary. Features found that drive pop-out include colour, luminance, movement and size. A soldier on watch would easily see a single person running across an open field within a military setting. However, detecting that same individual running among thousands of others in a crowded marathon would be slower and more difficult. A stimulus's degree of uniqueness across the visual field easily guides attention to it.

Not all search scenarios involve targets with a pop-out feature; therefore, parallel search is not the only mechanism searchers can use. Recalling the scenario of an individual runner in a crowded marathon, the individual would not pop-out from the other runners, and an alternate to parallel search is required. When a parallel search is impossible, visual search becomes harder, slower, and requires a more deliberate serial search (Treisman & Gelade, 1980; Wolfe, 1994). Serial search involves deliberately focusing attention on a single object and determining if it is a target, then examining the next object, and so on. Factors that may lead to serial search are increased heterogeneity of distractors, smaller differences between targets and distractors and higher numbers of targets (Humphreys et al., 1989; Wolfe, 2007). These stimulus-related factors combine to determine if parallel or serial search is used and how fast and easy the visual search task is.

Other factors unrelated to the stimulus also make serial search difficult. Search termination refers to a searcher deciding that all targets have been found or that no target is present. This mechanism can also contribute to the difficulty or speed of the search. The Wolfe (1994) Guided Search model proposes that potential targets are sorted based on bottom-up and top-down influences on each object, referred to as an object's activation value. A searcher checks each object in sequence based on the degree of activation, only terminating the search when they determine the likelihood is virtually zero of a target being in the remaining objects or that the search has taken too long. If the searcher is unsure how many targets are present in a scene, this can make the decision to terminate the search more complex and uncertain. Other models propose different mechanisms but yield the same prediction nonetheless (e.g., Duncan & Humphreys, 1989).

In sum, serial search requires more effort than parallel search to select which features are used to sort/filter the objects to be searched, perform the sort/filter, attend to an object, track which objects have already been checked, and decide when to terminate. This effort likely consumes cognitive resources such as working memory, and depending on the search context, the resources available might be limited due to competing demands from other tasks. So, any time that parallel search is not possible and serial search is required, there will be an increased demand for cognitive resources and a reduced search performance.

Operational Search

Despite the visual search literature being commonly based upon studies involving simple stimuli (e.g., Most et al., 2001), it does provide a foundation to understand operational search. Visual search that takes place in applied professional settings (hereafter "operational search") includes many of the features that make visual search more complex (Speed, 2015). Generally, an operational search occurs within dynamic and visually complex scenes. Targets are not always easily identifiable or clearly defined in advance. Environmental factors such as noise, distraction and competing tasks can also contribute to overall cognitive demands. So not only does the primary search task become more complex and demanding of cognitive resources, but concurrent tasks and situational factors place greater demands on those same cognitive resources. For example, an aerial search conducted by a military aviator may occur under significant time pressure whilst communicating with team members and with costly consequences for errors.

Within operational environments, the target is typically low salience relative to the distractors. In a military context, camouflage is commonly used to hide people and equipment, deliberately making targets less salient within the visual scene. Unlike laboratory-based tasks where the target is often obvious due to salience or clear instructions (e.g., "find the red circle"), in operational search, the target is less likely to include a high-salience distinguishing feature known in advance. Thus, operational search is unlikely to be visually simple enough to engage rapid parallel search and instead typically presents a situation that requires slow, inefficient and effortful serial search. Therefore, these difficult operational search contexts allow AR cues to aid operators by providing highly salient visual cues overlaid on potential targets. These cues may be sufficiently distinct, resulting in pop-out and rapid direction of attention via parallel search mechanisms. In principle, this should free up cognitive resources for searching and other concurrent tasks.

Inattentional Blindness

One of the plausible costs that may be associated with AR-based cueing is inattentional blindness. Within laboratory and operational search contexts, inattentional blindness is a common attentional failure. This attentional error occurs when an unexpected but otherwise plainly visible object fails to reach awareness (Mack & Rock, 1998; Simons & Chabris, 1999) and can occur despite overtly attending to the missed object (Hutchinson et al., 2023; Richards et al., 2012).

Initial studies of inattentional blindness focused on simple vigilance tasks involving geometric stimuli (see Mack & Rock, 1998 for a review). Subsequently, this phenomenon has also been seen in settings beyond the laboratory (Chabris et al., 2011) and during applied tasks involving Augmented Reality (Dixon et al., 2013). The extent of inattentional blindness is determined in a given task or visual scene by features including the inherent cognitive load, how unexpected an object is, and the homogeneity of stimuli within that scene (Mack & Rock, 1998).

Inattentional blindness is reliably replicated when an unexpected object is presented within the visual scene while participants perform a cognitively demanding task. For example, Hughes-Hallett et al. (2015) found clear evidence of the impact of cognitive load. Their study had a high-load condition where participants counted instrument movements in a video of a surgical procedure and a low-load condition where participants just watched the video with no counting task. Those in the high-load condition showed significantly more inattentional blindness than those in the low-load condition.

One expected benefit of AR cueing is reduced cognitive load from simplifying target search, which should translate to decreased inattentional blindness. However, when Yeh and Wickens (2001) introduced cueing into a military aerial search task, they found cueing increased blindness to critical, uncued, high-value targets. Importantly, even when participants were explicitly primed to prioritise the search for these high-value targets, these targets remained unnoticed more often when cues were present, and additional cognitive resources were expected to be available for deployment to the search for these prioritised targets. Indicating a potential cost associated with cueing or other mechanisms may be at work beyond cognitive load alone.

Protocol

Experimental investigations of inattentional blindness commonly use a paradigm involving the trial sequence of training trial(s), critical trial, a divided attention trial, and a final full-attention trial. This method was pioneered by Mack and Rock (1998) and remains influential today. This protocol controls for *expectancy,* a key element defining inattentional blindness and also validates that the missed object was otherwise obvious – an important criterion for an inattentional blindness effect. The protocol typically involves participants performing a search task throughout all trials, such as counting shapes bouncing off the edges of the screen or basketball passes. During the initial training trials, participants only perform this task. The initial training trials serve two functions. They train the participant to perform the primary vigilance task (such as counting the bounces of shapes against the screen edge) whilst manipulating their expectations. During the training trials, only those stimuli related to the primary task are shown (often targets displayed amongst task-irrelevant distractor stimuli), so over successive trials participants likely develop confidence that the frequent, salient, task-related stimuli they are actively monitoring will remain the only stimuli present in subsequent trials. That is, successive identical training trials cumulatively invoke an expectation of continuity.

Then, in the critical trial, participants perform the same search task using the same stimuli, but unbeknownst to them, an unexpected but otherwise obvious object is introduced. Depending on the experiment design, participants indicate they have seen an unexpected object during the critical trial (e.g., via a keypress) or are questioned about any unexpected objects immediately after the trial. Based on participants' responses, participants are classified as *noticers* or *non-noticers*, with non-noticers considered to have experienced inattentional blindness.

Crucially, the critical trial marks a turning point. Before the critical trial, there was no indication of the possibility of an unexpected stimulus. After the critical trial, experimenters presume that participants are aware of the possibility of the unexpected object, because participants either noticed it directly or were explicitly questioned about its presence (thereby implying the unexpected stimulus' presence). Therefore, when the unexpected object is shown again on the subsequent trial, the experimenter cannot assume that the object's presence in entirely unexpected. The next trial, the "divided attention" trial, proceeds identically to the critical trial, but on this trial, it is assumed that participants divide their attention between the original task and searching for the previously unexpected objects. The divided attention trial acts as a control trial to obtain noticing rates when the object is no longer unexpected or is more expected than on the critical trial.

Some studies use a different control in conjunction with the divided attention trial or in its place. This second control condition is the full-attention trial (e.g., Koivisto et al., 2004). The primary search task is not performed at all during the full-attention trial.

Providing a measure as to whether the unexpected object was visible under ordinary, passive viewing conditions. This control disambiguates the circumstance in which a participant fails to detect the unexpected object on the critical trial. They may have experienced inattentional blindness due to concurrent search, or the unexpected stimulus may have simply been too small, dim or otherwise non-salient to have attracted attention. Alternatively, perhaps the display was simply too cluttered with icons or overlays to allow detection of the unexpected object. The full attention task resolves this potential confound, as for inattentional blindness to have occurred, the object must be detectable under the ordinary viewing conditions present in the full attention control.

Operational Contexts

Within the present applied research domain, managing expectations via training trials also increases external validity. Many operational search tasks involve long periods, potentially hours, of only frequent targets. For example, a SONAR operator on board a submarine often spends hours reviewing displays of known targets' signals. Yet this operator must remain vigilant for unexpected or low expectancy signals indicating a possible adversary. They cannot simply become fixated on reviewing expected or frequent targets. However, whenever there is a low expectancy target and a concurrent monitoring task, there is the potential for inattentional blindness.

The distinction between inattentional blindness and other errors is important in operational search. Within military contexts, enemy combatants often actively camouflage personnel and equipment. If a soldier attended to an object but subsequently failed to identify it as a threat, this would best be defined as a classification error and not inattentional blindness. Similarly, Memmert (2010) contrasts inattentional blindness with the attentional errors observed in experimental studies of misdirection (e.g., magic tricks) examined by Kuhn and colleagues (e.g., Kuhn et al., 2008). In Kuhn’s magic studies, participants actively search for the “trick” but fail to see the sleight of hand due to active deception and misdirection. The present experiment focuses exclusively on errors of inattention in ordinary (dynamic, visually complex) environments and the impact of visual cueing on these errors of inattention, not other factors such as active deception or classification errors.

Determinants

Beyond the previously discussed link between cognitive load and inattentional blindness, results from divided attention trials point to expectations as an important determinant of inattentional blindness (e.g., Koivisto et al., 2004; Mack & Rock, 1998; Simons & Chabris, 1999). Under these divided attention conditions, Koivisto et al. (2004) demonstrated the influence of expectations on inattentional blindness. In this task, people did not tend to observe the unexpected object while under high cognitive load during the critical trial. Still, they did observe the unexpected object (even under load) once it was implied that an unexpected object might be present in the scene (on the divided attention trial).

Inattentional blindness protocols that include divided attention trials tend to be one-shot experiments: people either notice the unexpected object on the critical trial or not. Once the critical trial has been completed, the participant has seen or been informed about the unexpected object, rendering it less “unexpected” and leading to people more reliably reporting its existence in subsequent trials. Other researchers have experimented with different protocols that allow for more measurement occasions. Yeh and Wickens (2001) demonstrated a sustained version of inattentional blindness. Although participants knew that uncued and infrequent (and thus low-expectancy) threats were possible, they often missed them. In this sense, these misses were classed as errors of inattentional blindness. Therefore, expectancy is not simply an all or none division, as the classic distinction between critical trials and divided attention trials implies. Participants can demonstrate inattentional blindness for low prevalence targets that they know may occur at some point and are important, but at any given moment, the expectancy of these targets is low. Searching for these possible but unlikely, high-value targets is common in applied settings: detecting a mechanical defect on an aeroplane, searching for contraband in baggage screening, and detecting abnormal growths on a medical scan (see Wolfe et al., 2013 for a review of prevalence effects).

Attentional sets are another contributor to inattentional blindness. Attentional sets can be considered a filter for attention where only stimuli that match the current filter setting will be attended whilst other stimuli will be ignored. The filter is typically based on a feature such as colour, shape or luminance. Most et al. (2005) found that dissimilarity between targets and the unexpected item leads to higher inattentional blindness due to the unexpected item not matching the attentional set currently engaged for the primary task and, therefore, being filtered out. They demonstrated the influence of attentional sets for shape and luminance using black and white circles and squares (and found a similar effect using face stimuli). Additionally, an attentional set was also seen in Simons and Chabris (1999), where those attending to black shirt basketball players were more likely to see the gorilla than those attending to white shirts, suggesting either a colour (black shirts/black gorilla) or luminance (dark shirts/dark gorilla) attentional set. In summary, objects that do not share characteristics of the primary search target object are less likely to be attended to.

Orienting, Attention and Awareness

So far in this chapter, the word "attention" has been used as a catch-all phrase to indicate orienting, attention and awareness, whether that be in relation to the visual search literature, operational search settings, or inattentional blindness. But more precise definitions will be required for this study. Physically orienting one's focus to an object can be considered overt attention. Research has also shown that another form of attention – covert attention – occurs when one attends to something without orienting to it (Posner, 1980). However, it can be difficult to separate overt from covert attention. Under normal circumstances, covert attention is a precursor that guides the saccade, resulting in overt attention to the target stimulus a moment later (Posner, 1980). Both forms of attention are often distinguished from a third construct: awareness. Quantifying awareness is a complex endeavour (see Newell & Shanks, 2014). Still, for our purposes, we consider awareness to be indicated by a stimulus being processed elaborately enough that it could, in principle, support a behavioural response (e.g., clicking on a target stimulus).

Richards et al. (2012) found eye-tracking evidence for the separation of covert attention, overt attention, and awareness. Participants performed an inattentional blindness task, with the primary task involving black and white Ts and Ls moving around a screen. The unexpected object was a red cross that moved horizontally across the screen, and participants were questioned after the trial if they saw anything other than the Ls or Ts. Eye-tracking indicated that not all participants who reported seeing the red-cross overtly fixated on the red-cross, and many who overtly fixated on the red-cross failed to report it. In sum, Richards et al. (2012) results suggest that visually orienting on a stimulus is neither necessary nor sufficient for awareness: an object can be detected without ever being fixated upon, and it can fail to be detected even if it is fixated upon.

Research such as Richards et al. (2012) infer covert attention due to participants reporting noticing the unexpected object but eye-tracking indicating they failed to overtly attend to it. Direct measurement of covert attention is difficult, however, one technique that can help is via Steady State Visually Evoked Potentials (SSVEP; e.g., Walter et al., 2012). SSVEP recordings are collected via electroencephalography (EEG), electrodes on the skull collect neurological responses to stimuli that are flashing in the visual field (Norcia et al., 2015). Brain responses oscillate at the same frequency that the stimuli are flashing at, due to a combination of the visual and cognitive processing of the stimuli. The amplitude of the oscillating signal received increases with increased attention to the stimuli and the relationship between signal strength and attention is considered robust (Norcia et al., 2015; Walter et al., 2012). Such a measure suggests an opportunity to measure covert attention directly rather than through inference, reducing the need to specifically target experiment design towards enabling such inferences.

Visual Cues

One application for AR is providing real-time cues that conform to the size and spatial location of the target object to improve visual search performance (Galster et al., 2016; Hancock et al., 2013). This application of AR therefore intersects the visual search, attention, and cueing literatures, as such cues improve search performance by rapidly directing attention to a target using bottom-up and top-down influences (Anderson et al., 2011). Existing visual search theories and concepts, such as pop-out, parallel vs serial search and attentional sets, can be applied to AR cueing research. Once a visual cue has been projected onto a visual scene, it becomes an integral part of the visual input we perceive and process. However, as the cue is deliberately designed to stand out, it will have properties such as colour, luminance or depth that the rest of the scene does not. Applying existing visual search theories to these unique properties of cues should guide explanations of how and when cues work best and, equivalently, when their presence may induce costs on performance.

How Do Cues Work?

As presented via AR systems, visual cues provide information about important objects or events within the current visual scene. Cues can take various forms (Devagiri et al., 2022). Examples include highlighting an enemy combatant on a battlefield or a warning light indicating a potential system failure in an aircraft. However, the focus of this experiment is what is typically referred to as peripheral cues.

Peripheral cues are the most direct form of cueing as they rapidly direct attention to the specific location of the target without requiring interpretation (Most et al., 2005; Posner, 1980), usually due to some basic feature of the cue resulting in a pop-out effect, thereby directly attracting attention to the cued location. A soldier looking through night vision goggles and seeing a bright, high-contrast green rectangle displayed around a friendly soldier would be an example of a peripheral cue (see Woodward & Ruiz, 2023 for more examples). Such a cue would be highly salient against the complex visual scene presented by the night vision goggles and would pop-out. Peripheral cues tend to draw attention through bottom-up influences, where they are found via parallel search.

Although peripheral cues work predominantly through bottom-up influences, top-down influences also play a role. Prior experience is a top-down influence that can alter a cue's value, cost or priority. For example, previously important cues (because they predicted financial reward) continue to capture attention more effectively than equally salient but unimportant cues, even when those cues are no longer task-relevant (Anderson et al., 2011; Le Pelley et al., 2015). A further possible top-down influence is expectations which may drive an attentional set for some specific cue feature, e.g., colour. The attentional set will help filter out objects that do not match the set, improving performance for responding to the cue (Folk et al., 1992). Thus, bottom-up salience and top-down influences interact to guide attention towards visual cues.

Cues and Inattentional Blindness

Cueing targets by quickly directing attention to the cued target will result in faster response times. However, those same cues will also draw attention away from other uncued or low-expectancy targets resulting in potential blindness to those targets (e.g., Yeh & Wickens, 2001). Initially, cues will draw attention through the bottom-up salience of the cue. However, after several presentations, participants may become primed to the cues due to the benefits they provide. Over several trials, the initial top-down attraction to informative cues will begin to draw attention through learned or “derived” attention (Le Pelley et al., 2016), potentially influencing searches habitually or automatically. This priming may lead to selecting an attentional set based on the properties of the cue, further directing attention to the cues. The combination of bottom-up salience, attentional sets, learning mechanisms, and top-down informational value of AR cues ought to make it hard to divert attention towards under-specified, unexpected or infrequent events, thus leading to increased inattentional blindness.

Within the operational search literature, strongly attending to cued stimuli is often referred to as attentional tunnelling (Wickens & Alexander, 2009). In a context where there is only one search target that is reliably cued, tunnelling attention primarily towards cued objects may produce an optimal outcome. However, as discussed previously, operational search rarely occurs in simple environments, and attentional tunnelling is unlikely to produce optimal outcomes. In complex environments, operators need to strategically allocate attentional resources to search for cued and uncued targets, assuming the cueing system is imperfect.

Cue Reliability

It is rare for automated decision aids (such as an AR cue) to be perfectly reliable. This unreliability has elicited concerns about how well people use imperfect decision aids (Parasuraman & Riley, 1997). Within an operational search, the impact of an imperfect cueing system on overall operator performance is complex, as it simultaneously depends on other factors, including trust, context, cognitive loads, type of system errors, concurrent tasks, and expectations (Dixon et al., 2007; Wang et al., 2009; Wickens & Dixon, 2007; Wickens et al., 2006). Such inter-relationships have challenged quantifying the impact of cue reliability on performance.

Decision aids are often forced to make a decision based on probabilities derived from incomplete and inaccurate data in complex environments rather than a simple yes/no decision (Parasuraman & Riley, 1997). In an AR cueing system that is expected to choose between classifying an object as a target or non-target, there are two possible errors: false alarms and misses. A false alarm occurs when an object is incorrectly flagged as a target when it is not, whereas a miss is when an object is not flagged when it should have been. Reliability decreases as false alarms and/or misses increase. Reliability is commonly calculated using metrics that consider both error types (false alarms and misses), such as the sensitivity calculation for Signal Detection Theory (Green & Swets, 1974), and others that do not, such as the proportion/percentage of correct responses (Wickens & Dixon, 2007). Importantly, regardless of how the reliability of a decision aid is measured, reliability and types of errors generated cannot be ignored, as these two factors influence how decision aids are used (Dixon et al., 2007; Wickens & Dixon, 2007).

Logically there must be a point at which cueing (or any decision aid) is too unreliable to be of value, but that point appears to vary widely based on context. Wickens and Dixon (2007) conducted a literature review and concluded that a decision that was less than ~70% reliable was no better than unaided performance. Wang et al. (2009) found support for the ~70% reliability lower bound, but in contrast, Wickens et al. (2006) found that reliability as low as 60% improved performance. Comparing these two studies, Wang et al. (2009) used a single task with participants expected to identify friend from foe, whereas Wickens et al. (2006) used a dual-task paradigm where participants identified friend from foe whilst simultaneously monitoring the health of the simulated uncrewed aircraft they were controlling. The differences between these two studies suggest that any benefits from lower reliability automation are based on contextual factors, in this case, single task vs dual task demands. The findings make it difficult to justify generalising the ~70% reliability lower bound to all contexts, but it does suggest two principles. First, relatively high levels of reliability are expected by operators (all studies found minimal support for 30% reliable cues). Second, because the utility of an aid depends on its specific context, the overall performance of an operator using a decision aid should be compared against unaided performance within the same context.

In addition to the percentage/proportion of correctly cued targets, the *type* of error a system is prone to also influences how operators respond to cues. Dixon et al. (2007) showed that false-alarm prone systems are responded to differently than miss-prone systems. In their study, Dixon et al. (2007) kept the reliability rate consistent at 60% but varied if errors were due to misses or false alarms, finding that false alarms had a greater impact on overall performance than misses did. For the present research, cue reliability and types of cueing errors were essential considerations at the design and analysis stages to avoid inadvertently replicating prior reliability research or making ill-informed inferences from results.

Current Aims

The primary aims of the present research were to investigate if providing AR visual cues simplified and improved search performance for cued threats at the expense of poorer search performance for uncued threats. The specific context was a simulated military overwatch by a dismounted soldier.

Experiment 1 commenced this investigation by conducting within-subjects comparisons of performance in the presence of cues with perfect validity and reliability against performance without such cues. While a perfectly valid and reliable visual cue remains elusive in the applied domain modelled, this constitutes the strongest test of the influence of cueing, and so was chosen for the initial experiment. As discussed previously, cue reliability interacts with context specific factors such as task-load to influence how operators use the cues provided. Selecting a 100% reliable cue eliminated potential cue reliability confounds. The initial experiment had several related aims, but the primary aims were threefold. First, we sought to determine whether our simulated battlefield scene was sufficiently complex to elicit inattentional blindness. Second, the experiment examined whether this inattentional blindness would occur across several low-frequency appearances of a task-relevant target (akin to Yeh & Wickens, 2001). The third aim was to quantify the influence of presenting helpful cueing (valid and reliable) upon inattentional blindness to an uncued but important target.

To anticipate, Experiment 1 revealed empirical challenges when measuring inattentional blindness across repeated presentations and when participants experienced both cued and uncued trials—hampering quantifying the impact of cueing on inattentional blindness. The comparison between the validly cued and the uncued control confounded the presence of salient visual cues with their utility. When cues were present, they were valid and useful, so it was difficult to determine whether the cue's visual salience or utility attracted participants' attention. Experiment 2 addressed these shortcomings by presenting the unexpected element only once per trial and introduced an uninformative cue condition. This uninformative cue was only ~33% reliable, providing a condition where the cue remained visually salient but with no (or very low) utility. Finally, in Experiment 2, we introduced a novel measure of covert attention using EEG, referred to as steady state visually evoked potentials.

Experiment 3 replicated Experiment 2 as closely as practical using a virtual reality headset rather than the prior studies' screen-based display. The virtual reality experience was expected to improve perceived spatial realism for participants via a more immersive visual experience (Kisker et al., 2019) and a closer approximation to a real-world battlefield experience for participants. For example, AR cues on the battlefield are commonly provided via head-mounted displays (Woodward & Ruiz, 2023); the virtual reality headset would simulate this experience more than a screen-based experiment. In summary, the experiment was still a video game simulation of an overwatch. Still, virtual reality was expected to improve participant's spatial awareness, and the headset would further increase ecological validity compared to a screen-based interface.

Chapter 2: Experiment 1

Introduction

To reiterate, the present research focuses on investigating the costs and benefits of providing AR visual cues to aid operators in performing operational visual search tasks. Drawing upon inattentional blindness research and protocols, the present experiment will seek to determine if providing AR visual cues to aid in the search for one class of threat will hinder the search for uncued threats. Furthermore, if the provision of cues alters participant's global search strategies, i.e., how participants will search the scene.

Many factors within complex and dynamic operational contexts make visual search more difficult (Speed, 2015). The specific real-world implications of failing to detect a valid target vary with the operational search context, and the magnitude of damages can vary from zero to catastrophic. The present research will investigate the effects of providing AR visual cues within the specific operational search context of a dismounted soldier performing overwatch duty.

A dismounted soldier on overwatch duty is an example of a complex operational search environment. An overwatch involves a soldier taking a position and searching for potential threats in their allocated zone. Additionally, soldiers have demands beyond the allocated task, competing for attentional resources such as maintaining situational awareness. Situational awareness for a dismounted soldier might involve monitoring radio communications, listening for vehicle movements, and tracking friendly soldier movements. Their level of situational awareness partly determines a soldier's survivability and effectiveness (Skinner, 2019). These two competing tasks (overwatch and situational awareness) are both high priority, and a soldier must balance their available cognitive capacity between them. For example, failing at their overwatch search task by missing a potential threat removes the opportunity to classify and suitably action that potential threat.

The harm incurred from that missed opportunity will depend on whether or not the threat was genuinely hostile.

The complexity of operational search can lead to slower detection for some or all task-relevant stimuli. Delayed detection potentially leads to the same costs incurred for a missed detection. Even if the threat is eventually detected after a delay, the results could still be catastrophic. Boyd's OODA loop (Observation, Orientation, Decision, Action; as presented in Boyd & Hammond, 2018) provides insight into the importance of rapidly detecting potential threats in military situations. For success, you must process your OODA loop faster than your enemy does theirs (Endsley & Jones, 1997). Delayed attention to potential threats would slow your OODA loop, potentially benefiting your adversary and increasing the risk of catastrophic outcomes.

Cognitive Load

Cognitive load is a central consideration in this research for two reasons. Firstly, it has been shown to modulate inattentional blindness (e.g., Hughes-Hallett et al., 2015; Simons & Chabris, 1999). Secondly, reduced cognitive load is an expected benefit of AR cues due to reduced task difficulty. The cognitive load modulating inattentional blindness conclusion derives from experiments where the inattentional blindness protocol included conditions with varying cognitive loads. For example, the famous Simons and Chabris (1999) study involving a basketball game measured inattentional blindness based on noticing a gorilla walking through the game and included conditions of varying cognitive load. In this study, the difficult condition involved counting the number and type of basketball passes, whilst the easy condition only counted the number of passes. Results showed that 55% experienced inattentional blindness in the difficult condition compared to 36% in the easy condition. While under concurrent load, people were less likely to notice the unexpected gorilla in the scene.

Augmented Reality Cues Should Help

If cognitive load increases inattentional blindness, interventions that reduce cognitive load should reduce inattentional blindness. A benefit of AR cues is that they aid the visual search for the cued stimuli (Eisma et al., 2020). A reasonable presumption would be that such assistance would unencumber and allow flexible redeployment of in-demand attentional resources. Resources that may increase people's ability to notice other infrequent or task-irrelevant elements in the scene. In the present research, I investigate the impact of AR visual cues on inattentional blindness. In particular, I investigate the impact of peripheral visual cues co-located with a search target, which uses visual salience (or "highlighting") to direct attention.

A common implementation of overlaid visual cues is placing a marker or icon over a potential threat in the display, thereby acting as a warning signal (Cooke, 2019; Geospatial World, 2014). In application of this idea to the present scenario – a dismounted soldier performing overwatch duty – red overlay icons were placed over characters in the scene who are suspected to be a threat. In theory, placing valid, timely visual cues of this kind over suspected enemy combatants may save soldiers time in searching for otherwise subtle cues that a character may be a concealed combatant, thereby allowing them more time and resources to search for uncued or unexpected threats and other competing demands such as maintaining situational awareness. In the specific context of inattentional blindness, facilitated (cued) visual search for frequent or expected targets is anticipated to reduce inattentional blindness by allowing more cognitive headroom to detect and respond to uncued or unexpected visual events.

However, it is also important to recognise that visual cues are also visual elements and may increase competition for attentional resources. Once overlayed onto the real-world display, AR cues become another object or stimulus in the visual scene that acts to guide the search. In this experiment's scenario, a soldier monitors the scene for potential threats in a sustained visual search task. Therefore, we can deploy parallel/serial visual search models (e.g., Treisman & Gelade, 1980; Wolfe, 2007) to anticipate how high-salience AR visual cues may alter behaviour.

To review Chapter 1, these models propose that highly salient stimuli lead to a fast parallel visual search where visual attributes of the stimuli (such as colour) drive attention. The larger the visual difference between the target and other stimuli, the more likely the object is to rapidly and automatically draw attention, an experience commonly referred to as pop-out (Humphreys et al., 1989; Wolfe, 2007). The alternative is a slower serial search where each stimulus is attended to in sequence, and a decision is made about whether they are of interest. Serial search increases cognitive load due to the number and complexity of steps involved in the process (see Chapter 1). AR visual cues aim to decrease the cognitive load for the primary search by causing the cued stimuli to pop out, activating fast, efficient, parallel search rather than a more cognitively demanding serial search.

Task-Relevance

Beyond visual attributes of a stimulus, task relevance has been demonstrated to lead to prioritising one stimulus over another, comparable to pop-out visual features being prioritised. Within operational search literature, this is commonly referred to as attentional tunnelling and has also been observed in AR studies. For example, Eisma et al. (2020) found that providing augmented feedback resulted in attentional tunnelling to the provided augmented feedback. Specifically, participants conducted a simulated air traffic control task and needed to determine whether aircraft would converge. In one condition, they supplied operators with augmented feedback assisting them in judging convergence, with no augmented feedback in the second condition. Unlike AR visual cues, the feedback provided in this study was primarily designed to be informative rather than visually salient.

Participants' performance in the convergence judgment task improved for those receiving augmented feedback while reporting that the task was easier. Despite a lack of visually salient features, tunnelling of attention towards the augmented feedback was evident. The task relevance of the augmented feedback was sufficient to encourage attentional tunnelling.

Attentional tunnelling is also observed when participants are explicitly primed to search for specific uncued targets. Yeh and Wickens (2001) showed that attentional tunnelling towards task-relevant highlights led to a failure to redeploy residual cognitive capacity to uncued targets/events. Their study simulated a remote operator piloting an uncrewed aerial vehicle searching for ground-based threats. One threat, a missile, was never cued, but participants were informed it was the highest priority threat and search for it should take precedence over all other threats. Despite this priming, participants were 13% less likely to notice the missile when it appeared near a cued target than when it did not. The ability of the cues to attract attention overpowered the priming that instructed participants to search for the uncued high-priority threat. That is, cueing hindered the search for uncued, important events.

The Dual Search Task Problem

The obstacle in operational search contexts with AR visual cues is that the apparent benefits of a fast parallel search or attentional tunnelling to task-relevant cues may not produce an optimal outcome for all targets. If the cues provided were 100% reliable in cueing all target classes, then prioritising attention to them would produce an optimal outcome. However, recalling the previous discussion on cue reliability (Chapter 1), the systems underlying the AR visual cues are unlikely 100% reliable or may specialise in cueing only one target class. Why all targets may not be reliably cued is less relevant in the present research, the important consideration is that there is now likely to be a dual search task; a search for cued targets and a search for uncued targets.

The existence of a dual search task then raises the question of how will the costs and benefits from AR visual cues balance out. Put simply, will simplifying the search for one class of target (cued targets) allow the flexible redeployment of any now free cognitive resources in aid of the search for the other class (uncued targets), or will people tunnel their attention to the cues, unable or unwilling to redirect attention away from the salient and useful cues.

Global Gaze Behaviour

Eye-tracking provides an opportunity to investigate the impact of the presence of AR cues in a granular, time-varying manner. Because there are many ways of using gaze to quantify overt attention (patterns/ratios of saccades, fixations, and scanpaths) there is an additional opportunity to use each of these metrics in combination. This combined approach may be particularly impactful when seeking to quantify attention in a complex, dynamic scene with several search targets, multiple varied distractors and, in some instances, additionally overlaid cueing stimuli.

For example, consider what might occur if AR cues induce "pop out", i.e., preattentive capture. People may locate the cues rapidly (preattentively), as indicated by more frequent and faster overt attention shifts between cued targets (Duchowski, 2017). This should be indexed by more efficient saccade behaviour, whereby people would tend to saccade directly between cued targets rather than serially searching through spatially adjacent objects (including both targets and distractors). Such an explanation would, in isolation, suggest that cueing should lead to shorter scanpaths and a greater proportion of dwell time on cued targets, than on uncued stimuli. This is the conventional interpretation of scanpath lengths (e.g., Gilchrist, 2012).

Yet, in the present task there are multiple search targets, and both targets and distractors dynamically shift position. Most empirical work on gaze scanpath does not take place in scenes as complex as ours, and this has implications for the appropriate interpretation of typical gaze metrics, like scanpath. For example, on the assumption that cues facilitate rapid spatial search, it is also possible that people rapidly orient to a cued search target (target A), only to similarly rapidly move on to the next potential target (target B). Accurate performance on our task requires participants to constantly monitor for the next search target, so if their search was facilitated by the presence of cues, then perhaps they would perform more efficient saccades over longer distances and between sequential targets. That is, once they acquired a target, they may not merely rest their gaze point at that target, as is commonly the case in studies quantifying search efficiency using scanpath (Gilchrist, 2012). Under this logic, cueing would tend to facilitate search by generating longer scanpaths, not shorter ones; that is, the opposite prediction to that offered in the preceding paragraph.

The only way to discriminate between these two directional predictions is to consider other metrics of gaze behaviour. In this instance, if longer scanpaths are indicative of more efficient search, we should see more breadth in the coverage of the screen by sequential gaze points. By contrast, if longer scanpaths indicated slower, less efficient search, then people should have both only a narrow coverage of the screen and long scanpaths. The key point here is that it is the joint inspection of both metrics that provides insight as to the impact of cues on search behaviour; any lone metric is likely uninterpretable or ambiguous in this complex search scene. For this reason, we quantified dwell time for each class of target, scanpath, time to first fixation for each class of target, and search breadth. The precise calculation details are described in the sections below, but the intention was to adequately quantify the inherent richness of complex gaze behaviour generated to perform dynamic search in a complex scene.

External and Internal Validity

In applied use cases, the missed stimuli are inherently task-relevant; otherwise, failing to notice them would not be classified as a miss. Yet, in many lab-based inattentional blindness studies, the missed object is unusual or potentially noteworthy but is not inherently task-relevant (Memmert, 2010). This difference in the task-relevance status of the missed stimulus potentially limits the generalisability of lab-based studies to meaningful applied contexts.

Aside from the potential challenge to external validity, this observation also challenges the internal validity of inattentional blindness tasks. It is not inherently erroneous for someone to "miss" a task-irrelevant stimulus, however odd that stimulus might be. This is because optimal performance in the instructed task is achieved by remaining focused on their primary task to the exclusion of all else. So the failure to respond or report may not be appropriately characterised as an error (see Memmert (2010) for a related critique). For example, Simons and Chabris (1999) used counting basketball passes as the primary task, with inattentional blindness measured as whether you notice someone in a gorilla suit. However, noticing (or missing) a gorilla is irrelevant to the counting task, and participants were only instructed to perform a counting task. Similarly, in Hughes-Hallett et al. (2015) inattentional blindness experiment, participants counted surgical movements in the high-load condition vs only watching a surgical video in the low-load condition. Within this video were several surgical complications, and failing to notice those complications constituted evidence of inattentional blindness. Yet, in a strict interpretation of the instructed counting task, it is arguable that noticing surgical complications was irrelevant to the primary counting task.

This strict interpretation of task instructions stands in direct contrast to a surgeon performing real-world surgery. In this real-world application, the surgeon would know that detecting a potential complication was a vital component of their current duties, even if their immediate responsibility was, for example, monitoring the position of an internal sensor. The presumption is that participants in lab-based tasks would likewise understand that they had a baseline duty to notice and report highly unusual, unexpected, or potentially dangerous events observed, such as a gorilla in a university corridor. The full-attention condition controls for this assumption to some extent, as participants are not directly tasked to watch for an unusual event. Yet, they report it with high reliability. However, it is equally important to note that this control is imperfect; the presence of something unusual is typically indirectly implied by the time the full-attention trial occurs.

These specific concerns raise the broad issue that task relevance is insufficient to negate inattentional blindness (Hughes-Hallett et al., 2015; Yeh & Wickens, 2001). Nevertheless, wherever the unexpected object is task-irrelevant, a potential limitation to generalisability remains.

Experiment Overview

This experiment investigated the impact of AR visual cues on cued and uncued threats within a dismounted soldier overwatch scenario. The design was adapted from a common inattentional blindness paradigm (Mack & Rock, 1998) into a novel military scenario modelled in a custom video game. Participants performed 50% of trials with AR visual cues and the other 50% with no such aid. The participants were tasked to protect their four colleagues, depicted as four friendly soldiers onscreen. The participant achieved their goal by searching for and clicking on any threats in the scene. The search target ("threats") was defined directly and explicitly as any person wielding a weapon.

Unbeknownst to the participant, there were two types of threats shown across the course of the experiment, rare threats and frequent threats. The search for frequent threats was analogous to the cognitively demanding primary search task in the prototypical inattentional blindness paradigm. Rare threats were akin to the unsignalled introduction of a secondary, unexpected stimulus (e.g., a surgical complication or a gorilla).

The overarching research question was to investigate if providing AR visual cues helps or hinders participants' performance. Specifically, I was interested in whether visual cues improved detection speed and accuracy for cued targets and whether this facilitation incurred additional costs for the uncued and unexpected rare threat events. As already identified, two contrasting hypotheses arise. From the perspective of AR reducing cognitive load, reliable, valid visual cues should afford participants more resources to detect unanticipated events, thereby reducing inattentional blindness errors. However, suppose valid visual cues reduce cognitive load by inducing participants to rely on parallel search (because the informative visual cues "pop out"). Uncued events (including rare threats) may be filtered out. This explanation would posit that cueing should result in slower detections or misses of unanticipated rare threats. Put another way, attentional tunnelling to cued threats might prevent the redeployment of any cognitive savings obtained from the valid cues to the search for uncued threats.

Hypotheses

- Rare threat detection performance was predicted to reduce during cued trials compared to uncued trials, with cued trials yielding more misses and slower response times.
- Overt attention towards rare threats measured using eye-tracking was predicted to decrease during cued trials compared to uncued trials, with smaller dwell proportions to rare threats and later initial fixations.
- Frequent threat detection performance was predicted to improve during cued trials compared to uncued trials, with higher hit rates and lower response times.

- Overt attention towards frequent threats measured using eye-tracking was predicted to increase during cued trials compared to uncued trials, with larger dwell proportions and faster times to first fixation to frequent threats.
- Overt attention towards distractor characters measured using eye-tracking was predicted to decrease during cued trials compared to uncued trials, with smaller dwell proportions and slower times to first fixation during cued trials.
- The presence of cues was predicted to disrupt gross eye-movement metrics in a manner consistent with a higher gaze priority for cued than uncued threats.

Method

Participants

Forty-six participants from the Flinders University First Year Psychology undergraduate participant pool (34 female, 12 male; mean age 22.7 years, *SD* = 5.37) participated in exchange for course credit. All participants gave informed consent and during pre-screening reported normal or corrected-to-normal vision and normal colour vision. This research complied with the tenets of the Declaration of Helsinki and was approved by the Social and Behavioural Research Ethics Committee at Flinders University (Project ID: 2569).

COVID-19 related restrictions in place throughout the experimental period largely dictated access to participants and therefore final participant numbers. A priori power analysis conducted using G*Power3.1 (Faul et al., 2009) for a paired samples t-test comparing means using a two-tailed test, $d = 0.54$, $\alpha = .05$, and Power = 0.80 indicated a sample size of $n = 29$. The effect size was determined using the mean effect size of all studies included in the Hutchinson et al. (2022) inattentional blindness meta-analysis. This analysis was performed on the primary outcome measure analysis of rare threat hit rate. The results indicated the sample size was large enough.

Stimulus and Materials

The experiment was developed using the Unity (Version 2019.2.1.f.1, *Unity*, 2018) development platform in conjunction with the Unity Experiment Framework (Brookes et al., 2020). Participants sat approximately 60cm away from a Dell 25" 1080p Alienware (model #AW2518HF) monitor (Dell Inc., 2020) running at 120 Hz. Participants' head position was not fixed but measured at 60cm during the calibration before each trial. A Tobii 4C 90Hz eye tracker (Tobii AB, 2020) monitored eye movements.

The experiment was a video game representing a military overwatch scenario (Figure 1). The participant looked down on a market square where various characters moved around the scene performing multiple interactions and animations. The video game was a two-dimensional (2D) projection of a three-dimensional (3D) game scene with the Unity system (*Unity*, 2020) using two coordinate structures to reference the position of objects within the game scene. Unity uses a world-space referencing system to map the 3D model of the video game scene with objects such as buildings and characters positioned within the game scene using (x, y, z) references relative to the Unity world-origin (x=0, y=0, z=0). The Unity system dynamically maps these 3D world-space coordinates onto the 2D referencing needed to present the relevant objects on the display monitor and to log the positions of pre-determined visual objects. Our task logged the position of all characters in the scene, thereby allowing later gaze analyses.

Thirty-three characters of varying types were created using Adobe Fuse (Adobe, 2020) and Mixamo.com (IBM, 2020). The characters visually varied by shirt colour (as summarised in Table 1). All characters wore black trousers and brown shoes, except the friendly soldier characters, who wore camouflage-patterned trousers.

Figure 1

Example of the Experiment During a Cued Trial.

Note. Contrast on image adjusted to improve visibility for publication.

Table 1

Character Types and Attributes.

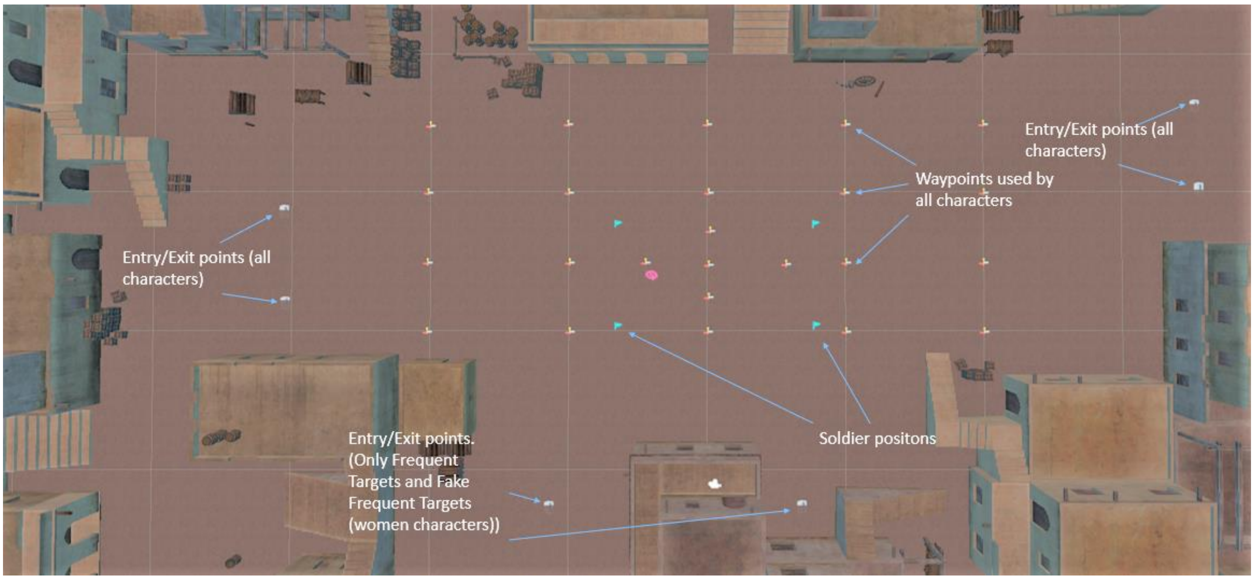
	Qty	Character Type	Shirt Colour(s)
	8	Frequent Threat	Red
	1	Rare Threat	Blue, Green, or Violet chosen randomly immediately before each screen onset
	4	Friendly Soldiers	Grey Camouflage
	20	Distractors	8 x Red, 4 x Green, 4 x Blue, 4 x Violet

Note. Contrast on images adjusted to improve visibility for publication.

Characters moved around the scene via various *waypoints* (Figure 2). A waypoint is a single (x, y, z) position in the 3D model of the scene that acts as a beacon for characters to move towards. Although Figure 2 shows the waypoints as coloured dots, waypoints were invisible in the game scene and, therefore, unknown to the participant. Which waypoints characters were assigned to traverse varied by character type (see Procedure). Each character attempted to move from one waypoint to the next in a straight line. The Unity "collision detection" system allowed each character to make small changes to its path to avoid collisions with other characters or obstacles. Consequently, in each scene the presence of multiple moving characters acted as an inherent randomization device. Characters were rarely able to follow a straight line between waypoints due to potential collisions, and so each character's path in each scene was largely non-determined even though their initial waypoint vectors were nominally predetermined. Participants could not merely memorise the paths of distractor characters and ignore those regions, as the motion depicted in any two scenes was never the same. Specifically, there were six possible waypoint networks for frequent threats & red shirt female distractors and 12 for the remaining 16 distractors.

Figure 2

Annotated Top Down View of Waypoints Used to Control Character Movement During All Trial-types in the Experiment.



Design

The experiment was a within-subjects design split into three blocks (training, critical, full-attention) with two trial types (uncued, cued) in each block. Cued trials provided the participant with a 100% reliable visual cue, presented as a red semi-transparent filled rectangle overlaid on the frequent threat characters, as seen in Figure 1. Uncued trials provided no such cues. No other characters were cued at any time. Training blocks always preceded the critical blocks, which preceded the full-attention blocks. The sequence of the two trial-types (cued, uncued) within each of the three blocks (training, critical, full-attention) was randomly chosen without replacement from the eight possible permutations of trial sequences. The design was based on a traditional inattentional blindness paradigm (Mack & Rock, 1998), with frequent threats equivalent to a distraction task and rare threats equal to critical stimuli.

Procedure

Onscreen and verbal instructions indicated that participants should look for and click on any threats to the friendly soldiers using the mouse (see Table 2 for the pertinent instructions per trial). Any character holding a weapon was defined as a threat. Participants were informed that in the cued trials, cues would only appear over frequent threats and during uncued trials, they would have no cues assisting threat identification. Participants indicated they understood the instructions and agreed to the collection of their data for research purposes by ticking the box at the bottom of the startup screen.

Table 2

Trial-Specific Instructions by Block and Trial Type

Block Type	Trial Type	Instructions
Training	Cued	This is a training trial where known threats WILL BE cued with a red rectangle around them. You will receive audio feedback if you miss a threat to help you learn the task.
Training	Uncued	This is a training trial where known threats WILL NOT be cued with a red rectangle around them. You will receive audio feedback if you miss a threat to help you learn the task.
Critical	Cued	During this trial known threats WILL BE cued with a red rectangle. From now on you will NOT receive any audio feedback if you miss a threat.
Critical	Uncued	During this trial known threats WILL NOT be cued with a red rectangle. From now on you will NOT receive any audio feedback if you miss a threat.
Full-attention	Cued	This trial is a little different. You will be aided by a colleague and you don't need to click on the red-shirt threats as your colleague will deal with those. Known threats WILL BE cued and your task is to keep watching and clicking on any threats you perceive.
Full-attention	Uncued	This trial is a little different. You will be aided by a colleague and you don't need to click on the red-shirt threats as your colleague will deal with those. Known threats WILL NOT be cued and your task is to keep watching and clicking on any threats you perceive.

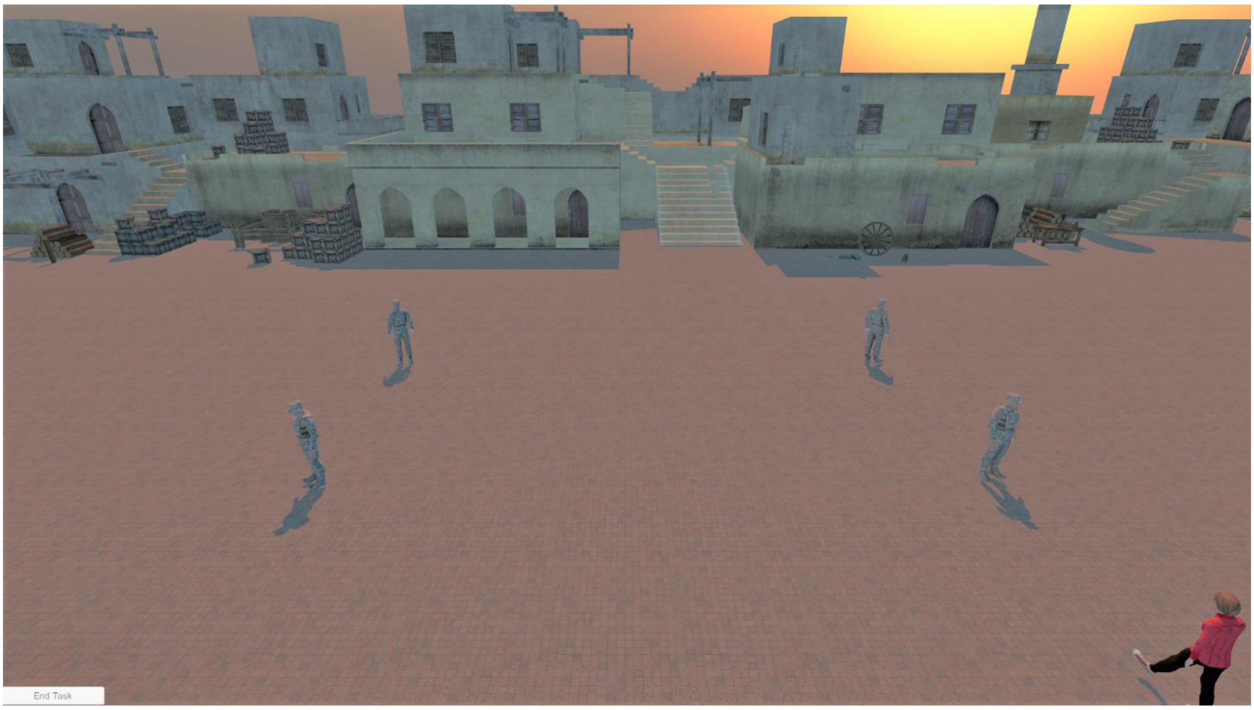
Training Trials

Before commencing the trial, participants were verbally provided with instructions for the upcoming trial (Table 2). After selecting the Begin Session button on the startup screen, the eye-tracker calibration commenced and followed the recommended Tobii procedure (*Calibration - Tobii Pro SDK documentation*, 2021). The game scene was then presented, showing the friendly soldier characters in a diamond formation around a central red square.

The experimenter then instructed participants to click on the red square when they were ready to commence the first trial. The trial began with the four friendly soldiers walking to one of the four soldier locations shown as blue flags in Figure 2 (and projected from the participant's point of view in Figure 3). The remaining characters entered the screen in a staggered fashion from the onset locations shown in Figure 2. Specifically, half had onset times staggered around 5 seconds after the start of the trial, whilst the other half entered the screen around 10 seconds after trial onset. Participants experienced all character types in this trial except the rare threat character.

Figure 3

Friendly Soldier Character Initial Locations.



Note. Soldiers moved intermittently and pseudo randomly between these four locations throughout the trial but always started in these four locations.

Classes of characters were assigned different movement patterns and interactions during the trial. Friendly soldier characters were always onscreen and positioned at one of the four locations in Figure 3. Friendly soldiers would move to the location either above or below their current location, with a delay of no less than 30 seconds. Therefore, at any given moment, zero, one or two friendly soldiers were positioned at any of the four soldier locations. A frequent threat attack occurred when one of the frequent threat characters intersected with a friendly soldier character. When this occurred, the frequent threat character would enter an animation state which involved them drawing a knife and attacking the friendly soldier character. It was important for friendly soldier characters to be static during these attack events. So, the experiment constrained the friendly soldiers' pseudorandom movements such that they would not move immediately before or during an attack event, nor could an attack commence during a friendly soldier movement. Participants were naïve to these constraints.

Frequent threat characters continuously traversed the scene. Depending on the trial sequence participants were assigned, frequent threats would either be 100% validly cued or not cued at all. To minimise the predictability of the target characters' movements, before onset, each frequent threat character was randomly assigned to approach the nearest friendly soldier character directly or to approach after traversing between one and four waypoints in a randomly assigned waypoint-network. Once a frequent threat approached the friendly soldier, they would commence an animation lasting approximately two seconds (M = 2.02 sec, SD = 0.04 sec). If the threat was clicked, the animation immediately terminated. They would then exit the scene from the left, right, or lower screen edge. Subsequent re-entry to the scene was delayed for 1-second after the last offset. In the training trials, an audible ding was heard after clicking on a frequent threat character and a warning noise if a frequent threat was not clicked on when required. These attacks occurred approximately 57 times over each trial, at a rate of approximately one incidence every 4 seconds.

All other characters in the scene functioned as distractors for the primary visual search task. There were three classes of distractor characters, male, red-shirt male and red-shirt female distractors. The four red-shirt female distractor characters acted the same way as the frequent threat characters by traversing the same waypoint network and intermittently walking directly towards the friendly soldier characters. The difference is that when these distractors approached a friendly soldier, they would *not* perform the attack animation. They only stood beside the friendly soldier for two seconds before moving away. Clicking on distractor characters had no effect and was not recorded.

The only difference between the eight male distractor and four red-shirt male distractor characters was shirt colour. They would traverse the scene using waypoint-networks randomly assigned before each onset. There was a 50% chance the characters would stop and perform a randomly selected animation at each waypoint to create diverse character movements in the scene. The possible animations included: arm stretching, looking away, looking around slowly, looking around fast, neck stretching, checking the time, searching pockets, yawning & stretching and three different waving motions. They would offset the screen to the left or right based on their assigned waypoint-network. Characters repeated the above process after waiting for 1-second and being given a new waypoint-network. Male distractor characters never approached a friendly soldier.

Trials lasted approximately two minutes (M = 115.11 sec, SD = 16.25 sec), ending when participants had successfully clicked on 20 frequent threats. The time per trial was non-deterministic due to variations in participants' detection rates and the wayfinding logic of the depicted characters. Participants hit and miss totals were displayed onscreen after the trial, and participants then completed the A-SWAT cognitive load questionnaire (see Measures section). The second trial in this block was similar to the first but started at the eye-tracker calibration step. Cueing was toggled off/on based on the assigned permutation of trial types. These two trials comprised the training trial block.

Critical Trials

The two critical trials occurred immediately after the second training trial. The participant was informed that no noise would be heard for subsequent trials if they missed clicking on a frequent threat. Otherwise, no indication was given to participants that this trial differed from the previous one. The unannounced changes in this block compared to the training block relate to the appearance of the rare threat and the trial termination logic. During these trials, the rare threat character would appear three times. Each appearance was randomly selected to be from the left or right of the screen, and occurred after a randomly selected delay range (30-60 seconds) after either the trial start or the last offset for the rare threat character. After onset, the rare threat always traversed directly through the middle of the scene carrying a rifle. The rifle was visible from the moment the character entered the scene. If the participant clicked on the rare threat character, the rifle disappeared, an audible ding was heard, and the character continued on their path unarmed, now appearing similar to a distractor character. They would exit the scene from the opposite side to their onset location, having been visible for approximately 20 seconds (M = 19.54 sec, SD = 1.73 sec). Cues were present or absent for the trial based on the participant's assigned trial-type permutation. Each trial ended after the third offset of the rare threat character, resulting in random trial lengths of approximately 4.5 minutes (M = 264.99 sec, SD = 14.86 sec).

Full-Attention Trials

The two full-attention trials were similar to the critical trials except that the experiment automatically clicked on frequent threat characters at a random time 0-2 seconds after the attack event commenced. This automatic action was 100% reliable and negated participants' need to act on any frequent threat characters. Participants were, therefore, verbally instructed that the system would automatically click any frequent threats attacking the friendly soldiers, and they no longer needed to search for them. Experimenters also reminded participants to continue looking for any threats. Each trial ended when the third rare threat exited the screen resulting in random trial lengths of approximately 4.5 minutes (M = 265.92 sec, SD = 15.31 sec).

Post Experiment

After the last full-attention trial, participants completed the Naivety, Video Game Experience, and general feedback forms as documented in the Measures section below. Experimenters then provided a verbal and written debrief explaining the purpose of the study.

Measures

Behavioural Measures & Character Tracking. The software platform (Unity) logged all characters' current screen locations 60 times per second during all trials. Additionally, click time & location were recorded anytime a frequent or rare threat character were clicked on, forming the basis for these characters' response time and hit rate measures. Gaze processing also used this data to classify each fixation's target(s).

Because all characters constantly moved throughout the scene, character regions of interest (ROIs) were defined dynamically using Unity "world coordinates". These 3D ROIs were a box 2m high, 1m wide and 1m deep centred on the midpoint of each character's skeletal frame (approximately their waist). The Unity system dynamically converted these 3D ROIs into 2D screen coordinates, meaning each 2D ROI scaled with the character size and distance from the participants' vantage point. Character ROIs furthest from the participants' vantage point measured 0.92° horizontally 1.83° vertically. Rare threat characters were constrained to cross the scene's centre. At this point, character ROIs measured approximately 1.97 by 3.62°. Unity recorded the 2D and 3D ROI coordinates when logging a character's location.

Gaze Measures. Eye Tracking recorded participants' gaze points and pupil diameters continuously at a sampling rate of 90Hz. Fixation calculations used a simple velocity threshold (Salvucci & Goldberg, 2000). Any two or more sequential gaze points with an instantaneous velocity of less than 100°/sec (Salvucci & Goldberg, 2000) were combined and classified as a fixation. The location of a fixation was defined as the centroid of the constituent gaze points that made up that fixation. The summed total linear distance between the centroids of sequential fixations was used to calculate scanpath distances.

Scanpath and fixation gaze metrics were normalised at the trial level because trial durations (and thus the opportunity to gaze) varied from trial to trial within and across participants. For a given participant, the third and fourth trials might vary by 10 seconds in duration. This means that the raw scanpath and total fixation duration metrics would differ between trials, even if the behavioural pattern were identical. Hence, gaze metrics were normalised by dividing the number of fixations and the scanpath metrics by trial duration (in seconds). Scanpath distances were therefore scored as a non-directional velocity of *scan pixels per second*, and the number of fixations was reported as *fixations per second*.

After gaze behaviour was demarcated into fixations and saccades, the fixations were further classified according to each fixation's target(s). Fixations were assigned to characters when the centroid of a fixation lay within a character's 2D ROI at the time the fixation occurred (measured in individual screen frames). This method is conceptually aligned with techniques often used in the experimental psychology literature (e.g., Richards et al., 2012). This method also meant that a single fixation could be assigned to two or more ROIs if, during that fixation, multiple characters' ROIs were overlapping. Our visually cluttered scene included many such instances, and there was no reliable process to determine which overlapping ROIs were the intended fixation target. To quantify the extent of this potential methodological challenge, I examined every fixation during the critical trials. Within all critical trials, 26% of fixations were not assigned to a character, 38% to a single character, and 36% to two or more characters.

This list of ROI-classified fixations was used to calculate character-based gaze metrics. The fixation duration for each character was individually calculated by tallying all fixations to that character, performed separately for all 33 characters per trial. The result was then grouped and summed based on each character's type (e.g., across all four friendly soldier characters) and trial number and then normalised by trial duration. This process yielded a mean proportion dwell time (hereafter, "dwell") for each character type: rare threat, frequent threat, distractors, and friendly soldiers. Although each class of character had different numbers, the resulting dwell metrics were comparable across stimulus classes. This is because each constituent measurement was normalized per onscreen moment per character. Thus, the resulting aggregated values were also normalized relative to the frequency and duration of each character's class. Of particular note is that for the rare threat characters, the denominator was not the trial length but the time the rare threat character was onscreen whilst the gun they were carrying was visible. These character-type dwell measures were our primary eye-gaze metric.

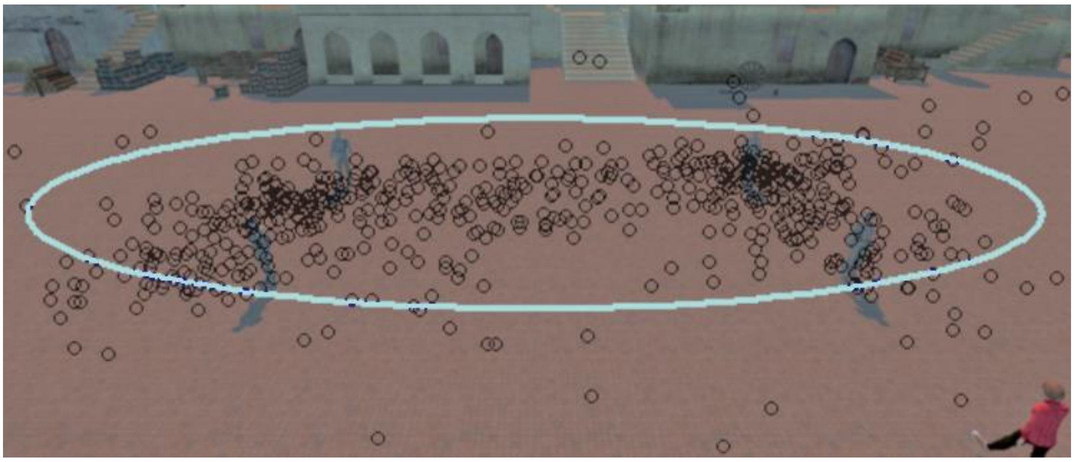
An additional character-based metric included was *time to first fixation* in seconds. This metric is commonly used within visual search studies (see Carter & Luke, 2020 for a review) as a measure of search efficacy or salience. Within this experiment, the metric enables a comparison of the effect of cuing on each character type's ability to attract attention. Each character's time to first fixation was grouped by character type and averaged per trial, resulting in a character-type mean time to first fixation.

I was additionally interested in exploratory analyses of the breadth of visual scene coverage that participants achieved with and without cueing, hereafter referred to as *search breadth*. Breadth was selected as the primary metric because the experiment was designed such that characters onset only from left and right of the display and only offset on the left, right, or bottom, thus compressing the vertical search axis. Various metrics can quantify how much scene width was gazed upon during trials, and I considered two.

The first was a "minimum ellipse" method representing the cloud of fixation locations per trial as an ellipse. Using code adapted from Duarte and Zatsiorsky (2002), an ellipse was calculated, enclosing 95% of fixations per trial in the smallest possible ellipse (see example in Figure 4). This analysis results in three parameters describing an ellipse, one per trial. The parameters were: width, height and angle. Only the width parameter was required for our search breadth analyses as it measured how broadly participants spread their gaze on the longest ellipse axis, which was consistently close to horizontal in orientation.

Figure 4

Example Plot of All Fixations for a Single Critical Trial Showing Calculated Ellipse Enclosing 95% of Fixations Overlayed on Opening Game Scene.



Note. Image for illustrative purposes only. Circles represent all fixations for a single trial with ellipse enclosing 95% of fixations. Overlayed onto an opening scene captured from a single trial.

The second metric (“distance to nearest friendly soldiers”) was a novel metric developed for the particular search task used in the present experiment. This measure exploited the observation that the task inherently pushed people's gaze towards the friendly soldiers (as that is where attack events commonly occurred). The metric was defined based on the ability to move gaze away from these high-likelihood search zones. The metric is conceptually aligned with similar measures in studies where fixation distance relative to task-relevant stimuli is recorded (e.g., Beanland & Pammer, 2010). Specifically, I defined the breadth of search as the mean of the Euclidean distances of each fixation centroid from the screen position of the nearest friendly soldier.

Greater values in either metric indicate more search breadth. However, they are conceptually different. The ellipse width measure can be considered more data-driven; it was ungrounded in the specific search task used in this experiment. Instead, the measurement was derived purely from the raw fixation points. In contrast, the distance to nearest friendly is grounded in the design elements of the specific task. Therefore, it only applies to studies using a similar search task in which visual stimuli elicit action-prompting events in a central, primary search zone.

Subjective Workload. Participants self-reported overall workload/difficulty at the end of each trial by completing the Alternative Subjective Workload Assessment Technique (A-SWAT, Luximon & Goonetilleke, 2001). This measure comprised three subscales (time load, mental effort, and stress load) in which participants marked the degree of each type of load on a separate horizontal line, with textual anchor points at the extreme left and right of the line. The anchor points for Time load were "Often have spare time. Interruptions or overlap among activities occur infrequently" on the left extreme and "Almost never have spare time. Interruptions or overlap among activities are very frequent" on the right. For Mental effort, the left anchor point was "Very little concentration required. Activity is almost automatic, and requires little attention", and the right was "Very high concentration required. Activity is complex and requires total attention". For Stress load, the left anchor point was "Task causes little confusion, frustration or anxiety", and the right was "Task causes intense stress due to confusion, frustration or anxiety". The A-SWAT total was the mean of the three sub-scale scores. Although this measure was collected as part of the current experiment, it is not the subject of this thesis and will not be reported or discussed within this thesis.

Video Game Experience. After participants finished all trials, they completed a questionnaire loosely based on Green and Bavelier (2007) to capture their video game experience. Participants recorded their current (last 12 months) and past (before the last 12 months) video game-playing behaviour by indicating what types of games they played, estimated average hours per week, and the age at which they started playing. Although this measure was collected as part of the current experiment, it is not the subject of this thesis and will not be reported or discussed within this thesis.

Naivety to Experiment and General Feedback. After the experiment, participants were questioned on their noticing of rare threats, their naivety to inattentional blindness and any general feedback they may have with the following questions:

1. Did you notice any threats other than the men in red shirts attacking your soldiers?
2. Did you notice the men carrying guns?
3. Have you ever heard of inattentional blindness?
4. Have you seen the Gorilla in the basketball game study?
5. Any general feedback, thoughts or questions?
6. Did you associate this study with inattentional blindness?

Trust and Perceived Accuracy. Trust in the cue, perceived cue accuracy and perceived self-accuracy were measured after the completion of all trials using measures taken from Merritt (2011). This measure was primarily a manipulation check, as the cueing presented in this experiment was perfectly valid and reliable. Participants used a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) and self-reported their trust in the cues by responding to six statements such as "I trust the aid". The mean response was used to measure the trust in the cues provided. Participants were verbally informed that the cues were a form of automated decision aid.

Perceived accuracy of the cues was directly queried with the statement, "Out of all of the trials you completed with assistance from the automated aid, what percentage of the time did you think that the aid was correct?". Participants filled in the statement, "I think the aid was correct ___% of the time."

Self-perceived accuracy of the participant's performance on uncued trials was queried with the statement, "Out of all of the trials you completed without assistance from the automated aid, what percentage of the time did you think that you were correct?". Participants filled in the statement "I think I was correct ___% of the time."

Statistical Analysis

Raw data pre-processing was conducted using Matlab (*MATLAB*). Statistical analysis was conducted using JASP (JASP Team, 2023) and R (R Development Core Team, 2020). G-tests were performed using the Desctools R package (Andri et mult. al., 2022). Statistical interpretation used the effectsize package (Ben-Shachar et al., 2020), based on Cohen (1988) for Cohen’s D, Funder and Ozer (2020) for Cramer’s V and Jeffreys (1961) for Bayes factors, unless otherwise noted. An alpha level of .05 was used to determine the significance of all analyses.

When required posthoc analysis was performed using the Bonferroni adjustments option in the JASP software package. This package reports a p_{adjust} value rather than an unadjusted p value. For example, if there are three comparisons then p_{adjust} equals p * 3. If p_{adjust} goes above one, then the software reports the value as one. The desired alpha value remains unchanged when determining if a result is considered significant. This is in contrast to the more conventional method of dividing the desired alpha level by the number of contrasts, with the relevant p values remaining unchanged. Within this thesis the p_{adjust} values are reported in alignment with the software packages used.

Unless otherwise noted below, as part of the data cleaning and preparation stage screening was performed on all measures. Outliers for continuous measures were checked using the identify_outliers function from the rstatix packing in R (Kassambara, 2021). Investigation of outliers was performed to determine if they were feasible values for the individual measures and not a result of measurement, technical, or other errors. Those deemed to be feasible values were retained in the data set and any subsequent analyses.

The assumptions for all statistical tests were performed. Continuous data will be analysed using ANOVA tests, due to these tests being robust to modest violations of assumptions (Boneau, 1960; Cohen, 1988), only gross violations of assumptions will be noted and alternate tests applied. Categorical data were evaluated using the G-Test (also known as the log-likelihood ratio test), all expected cell counts were checked to ensure values greater than five (Agresti, 2019).

Unless otherwise noted, this general approach applies to all experiments in this thesis.

Results

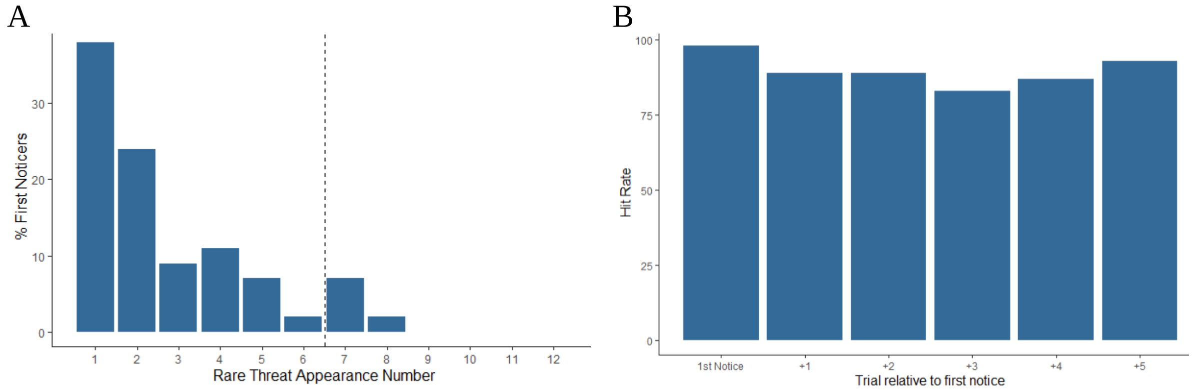
Behavioural Results

Rare Threat

The behavioural responses to the rare threat character appearances during the critical trial were the main measure of cueing-induced inattentional blindness. The rare threat appears twelve times for each participant. Figure 5A shows that by the last critical trial appearance, the majority of participants (85%) have noticed the rare threat at least once; with a high likelihood of continuing to notice once noticed for the first time (Figure 5B).

Figure 5

Rare Threat Noticing Patterns



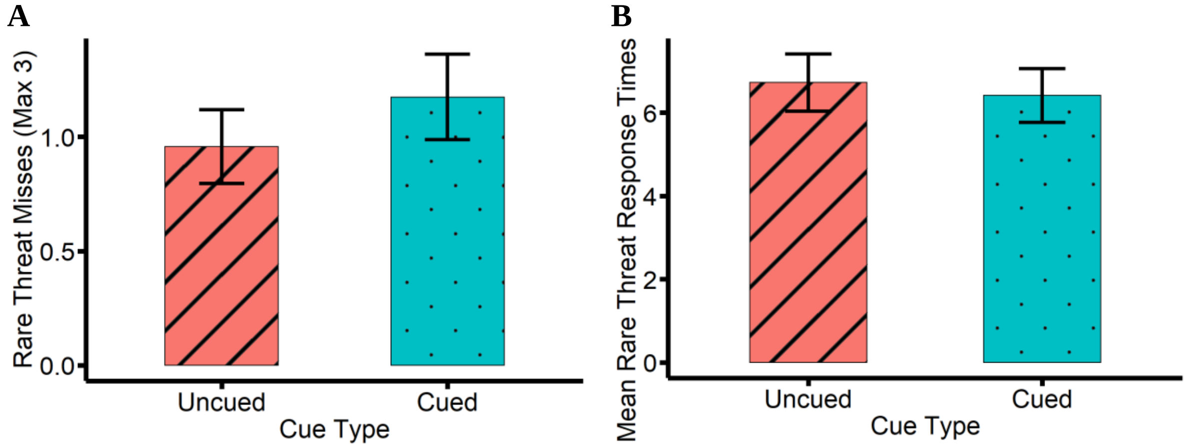
Note. (A) The dotted line indicates the end of the critical and start of the full-attention trials. (B) "First notice" on the x-axis indicates the first trial that participant noticed the rare threat, remaining columns showing the hit rates for subsequent trials. The "1st Notice" column represents all noticers and therefore always equals 100%.

Overall miss-rates and response times during the critical trials are shown in Figure 6. Rates of missing were analysed using a paired samples t-test comparing cued versus uncued critical trials. Response times were analysed similarly, but note that response times were only available when participants clicked on the rare threat; misses were not able to be entered into the response time analysis.

Analysis of the critical trials showed no significant effect of cueing and moderate evidence for the null hypotheses. This pattern of results was for both misses, $t(45) = 1.11$, $p = .274$, $d = 0.16$, $BF_{10} = 0.28$, and for response times, $t(31) = 0.35$, $p = .728$, $d = 0.06$, $BF_{10} = 0.20$. These results suggest that providing cues for the frequent threat characters leads to no change in inattentional blindness rates towards the uncued rare threat character nor any cost in the form of slower response times. The prediction that both response times and miss rates would be worse in cued trials was not supported.

Figure 6

Behavioural Responses to Rare Threat Characters by Cue Type During Both Critical Trials.

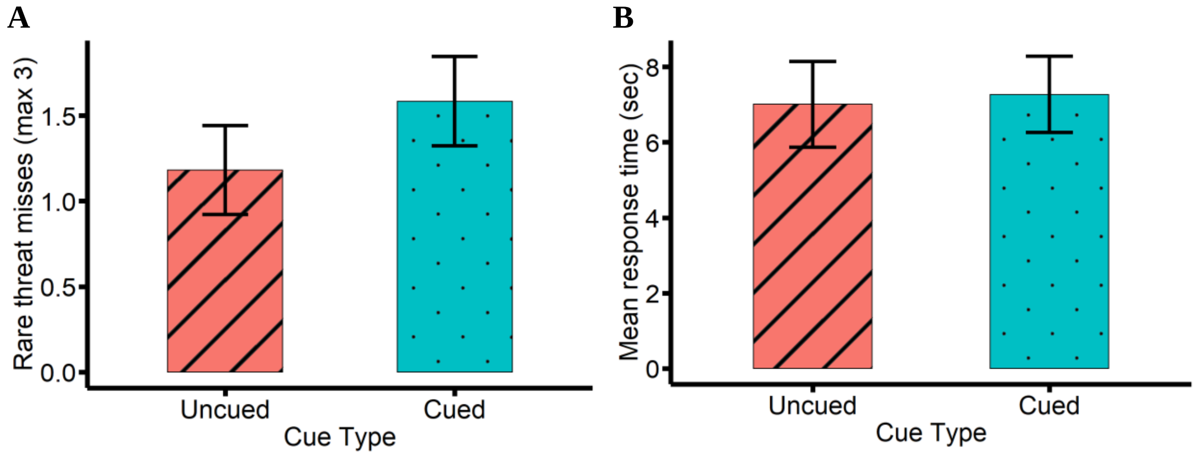


Note. (A) Rare threat miss count during critical trials, maximum misses of 3 per trial. (B) Rare threat response times in seconds during critical trials. Error bars represent standard error.

In understanding why no effect of cueing on inattentional blindness was observed, a further possibility was identified whereby people may respond differently on their second critical trial to their first. By the time they experienced their second critical trial, they already had three opportunities to see the rare threat. Under the conventional inattentional blindness protocol, our second critical trial might meaningfully be considered to be a divided attention trial. Consequently, I separately considered the first critical trial in isolation. Each person either received the cue or uncued critical trial first, so this meant using a between-subjects comparison rather than a within-subjects comparison, lowering statistical power (Cumming, 2014). The mean performance on the first critical trial is plotted in Figure 7. Results showed no significant effects and anecdotal evidence against differences between groups. This pattern was observed for both mean miss rate for the rare threat, $t(44) = -1.09$, $p = .284$, $d = 0.32$, $BF_{10} = 0.47$, and mean response times, $t(30) = -0.17$, $p = .866$, $d = 0.06$, $BF_{10} = 0.34$. This pattern of results matched the within-subjects comparisons above (Figure 6). Limiting analyses to the first critical trial did not reveal a significant effect of cueing.

Figure 7

Behavioural Responses to Rare Threat Characters by Cue Type Only During First Critical Trial (Trial 3).



Note. (A) Rare threat miss count during first critical trials, maximum misses of 3 per trial. (B) Rare threat response times in seconds during first critical trials. Error bars represent standard error.

Yet perhaps even this single-trial analysis may have been diluted by occasions in which the rare threat was expected, as distinct from the unexpected probe usually used in inattentional blindness studies. During traditional sustained inattentional blindness studies in which the unexpected stimulus is shown for a few seconds or longer, the unexpected stimulus typically appears only once (but see Dixon & Wickens, 2006; Pitts et al., 2012; Yeh & Wickens, 2001). To apply a more traditional inattentional blindness analysis to the present stimuli, the data were further filtered to include only the first appearance of the rare threat character (during Trial 3). Participants were then binary classified as non-noticers (i.e., experienced inattentional blindness) if they failed to click on the first appearance of the rare threat character, and noticers if they detected that rare threat. This resulted in 55% (n = 12) of those who experienced the uncued trial first and 71% (n = 17) of those who experienced the cued trial first being classified as non-noticers. Numerically more people missed the initial rare threat when it was presented in the context of visual cues. While the direction of the means was consistent with the primary hypothesis, this difference was not significant. A G-test (log-likelihood ratio test of independence) showed that the proportion of non-noticers between the cued and uncued trials was not significantly different $G(1) = 1.31$, $p = .252$, $V = 0.17$, $BF_{10} = 1.23$. Due to low noticer numbers (n = 17) for this first appearance, the rare threat response times were unable to be meaningfully analysed.

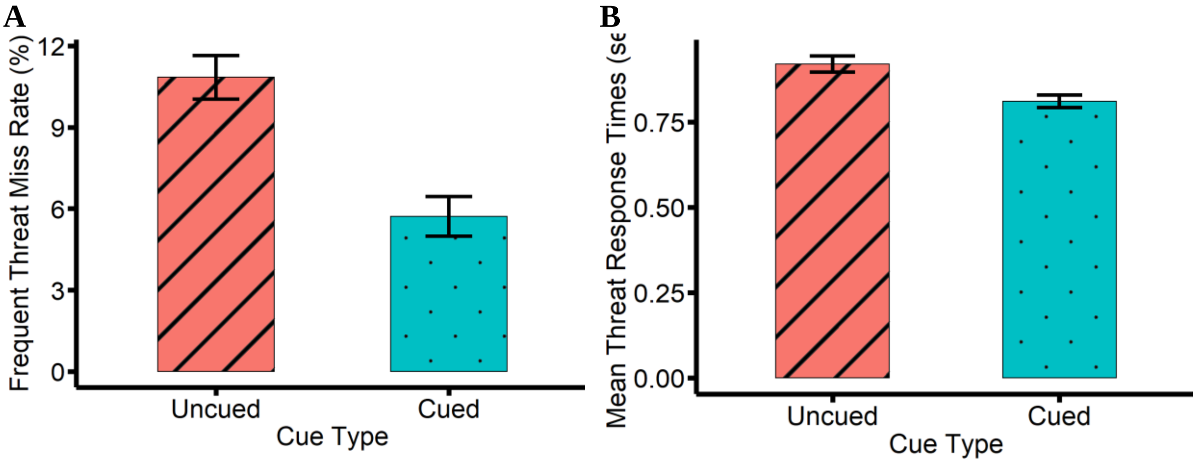
Despite investigating rare threat noticing rates and response times as a within-subjects comparison (both critical trials) and again as a between-subjects (first critical trial only) comparison, no evidence was observed that cueing had an effect. Inattentional blindness rates did not alter between the cued and uncued trials and neither did the response times for those that did notice the rare threat character. In summary, contrary to predictions cueing showed no significant effect on rare threat characters, however there was a trend towards the predicted relationships.

Frequent Threat

Frequent threat appearances performed two primary functions: they provided a cognitively demanding primary task and facilitated the manipulation of uncued vs cued trial types. I hypothesised that providing highly salient and task-relevant cues during the cued trials would improve all frequent threat behavioural measures. Figure 8 summarises the behavioural responses to frequent threats during the critical trials. Paired samples t-tests were used to analyse the effect of cueing on miss rates and response times for frequent threats. Significant facilitation for the cued group was seen on both measures: miss rates, $t(45) = 5.52$, $p < .001$, $d = 0.81$, $BF_{10} > 100$, and response times, $t(45) = 7.81$, $p < .001$, $d = 1.15$, $BF_{10} > 100$.

Figure 8

Behavioural Responses to Frequent Threat Characters During Critical Trials.



Note. (A) Frequent threat miss rate during first critical trials. (B) Frequent threat response times in seconds during first critical trials. Error bars represent standard error.

These large and unsurprising results support the idea that providing a visually salient and task-relevant cue will increase search performance for cued stimuli, and further speak to the experimental sensitivity of the design. That is, although the present task was relatively brief, its design was sensitive to detecting positive effects of cueing with high magnitude and confidence.

Eye Tracking

Exclusions

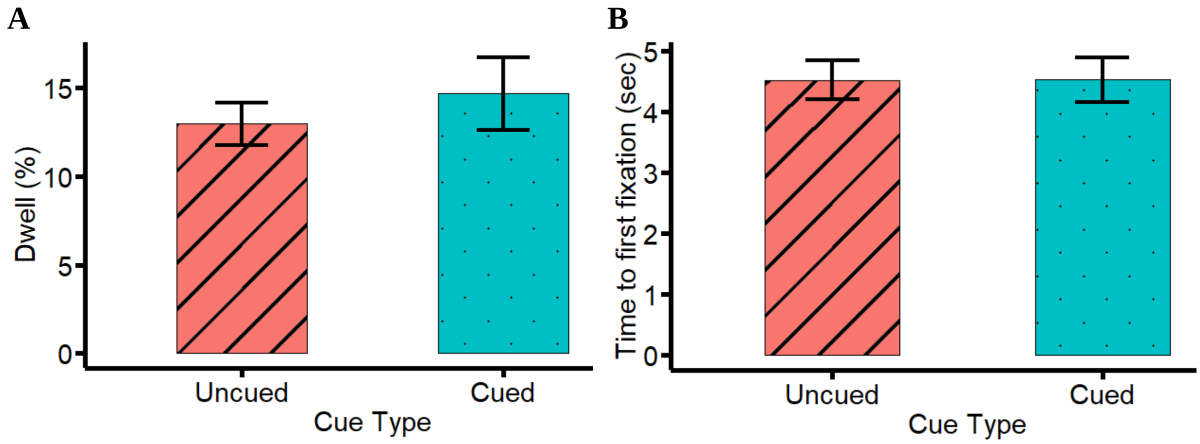
Participant #45 was excluded from all eye-tracking analyses due to a hardware fault resulting in few recorded fixations. Their total dwell time during critical trials (uncued trial: 22.13 secs, cued trial: 8.74 secs) was more than six standard deviations lower than the mean of all participants in both critical trials (uncued trials: M = 231.00, SD = 34.08, cued trials: M = 230.85, SD = 36.37).

Rare Threat

Eye-tracking measures were used to analyse differences in overt attention towards rare threat characters between uncued and cued trials during critical trials. A paired samples t-test was run for the effect of cueing on dwell (Figure 9A), showing no effect, $t(44) = -0.67$, $p = .506$, $d = 0.10$, $BF_{10} = 0.20$. Comparisons using paired samples t-tests showed no effect of cueing on time to first fixation (Figure 9B), $t(45) = 0.01$, $p = .994$, $d = 0.00$, $BF_{10} = 0.16$. The combined effect of these results showed that cueing frequent threat characters did not alter participants' overt attention to the uncued rare threat characters for either the first fixation after the onset of the character or in how long participants looked at the character whilst the gun was visible.

Figure 9

Eye Tracking Outcomes to Rare Threat Characters During Critical Trials.

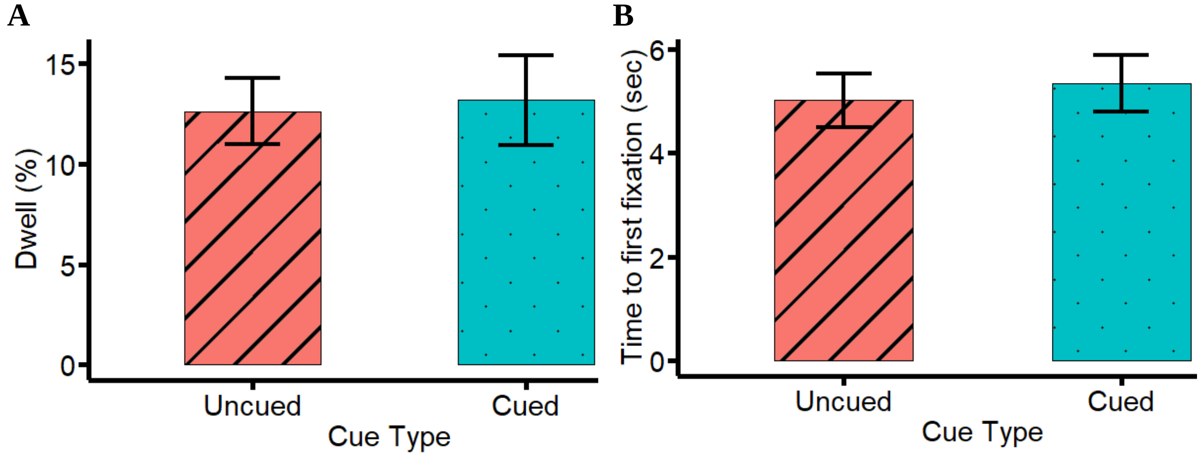


Note. (A) Rare threat dwell as a percentage of character onscreen time and gun was visible. (B) Rare threat time to first fixation in seconds. Error bars represent standard error.

Mirroring the earlier analysis of detection rates for the rare threat stimuli during only its first appearance, gaze metrics were also analysed during the first critical trial only (Trial 3). Independent samples t-tests showed no significant effect of cueing on either rare threat dwell (Figure 10A), $t(43) = 0.20$, $p = .841$, $d = 0.06$, $BF_{10} = 0.30$, or rare threat time to first fixation (Figure 10B), $t(43) = 0.43$, $p = .670$, $d = 0.13$, $BF_{10} = 0.32$.

Figure 10

Eye Tracking Outcomes to Rare Threat Characters During Trial Three Only (First Critical Trial).



Note. (A) Rare threat dwell as a percentage of character onscreen time and gun was visible. (B) Rare threat time to first fixation in seconds. Error bars represent standard error.

These results are comparable to the within-subject behavioural analyses above, which consider both critical trials. In sum, there was no evidence that trial order impacted on gaze towards the rare threat character.

Frequent Threats and Distractors

I predicted that participants' attention would be biased more towards frequent threat characters and less towards distractor characters during cued than uncued trials. Attention was measured using dwell and time to first fixation for the frequent threat characters and all distractor characters.

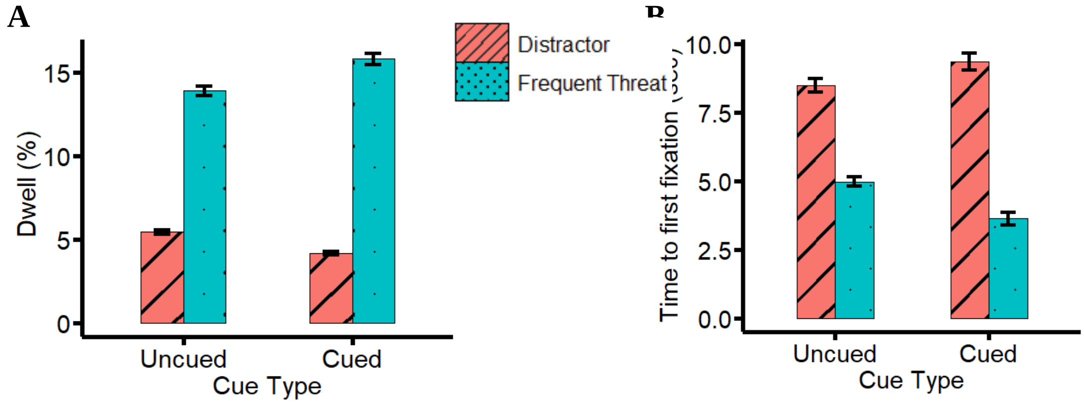
A two-way repeated measures ANOVA was performed to evaluate the effect of trial type (uncued/cued) and character type (distractor/frequent threat) on dwell (Figure 11A). The main effect for character type was significant, with extreme evidence that participants dwelled longer on frequent threats than distractors, $F(1, 44) = 1927.97$, $p < .001$, $\eta^2_G = 0.91$,

$BF_{10} > 100$. The main effect of cue type on dwell was not significant, $F(1, 44) = 1.76$, $p = .192$, $\eta^2_G = 0.01$, $BF_{10} = 0.31$.

The interaction between character and cue types on dwell was significant with extreme evidence, $F(1, 44) = 62.06$, $p < .001$, $\eta^2_G = 0.21$, $BF_{10} > 100$. Therefore, the effect of cueing was analysed for each character type individually. Comparisons using paired samples t-tests with Bonferroni corrections revealed that participants dwelt longer on frequent threats during cued trials than uncued trials, $t(44) = 6.17$, $p < .001$, $d = 1.22$, $BF_{10} > 100$, and showed the opposite pattern for distractor characters. Specifically, they dwelt less on distractor characters during cued trials than uncued trials, $t(44) = 4.16$, $p < .001$, $d = 0.82$, $BF_{10} > 100$.

Figure 11

Eye Tracking Outcomes for Frequent Threat and All Distractor Characters by Cue Type During Critical Trials.



Note. (A) Character type dwell as a percentage of trial length. (B) Character type time to first fixation in seconds. Distractors include all distractor types. Error bars represent standard error.

A two-way repeated measures ANOVA was performed to evaluate the effect of trial type (uncued/cued) and character type (distractor/frequent threat) on time to first fixation (Figure 11B). The main effect for character type was significant, with extreme evidence that participants were faster to fixate on frequent threats than distractors, $F(1, 44) = 331.77$, $p < .001$, $\eta^2_G = 0.66$, $BF_{10} > 100$. The main effect for cue type was nonsignificant, indicating no difference in time to first fixation between cue types, $F(1, 44) = 1.25$, $p = .271$, $\eta^2_G = 0.01$, $BF_{10} = 0.28$.

The interaction between character and cue types on time to first fixation was significant with extreme evidence, $F(1, 44) = 40.50$, $p < .001$, $\eta^2_G = 0.10$, $BF_{10} > 100$. Therefore, the effect of cueing was analysed for each character type. Comparisons using paired samples t-tests with Bonferroni corrections revealed a significant effect with extreme evidence of cueing on frequent threat character times to first fixation, with participants being faster to fixate on frequent threat characters during cued trials than uncued trials, $t(44) = 4.96$, $p < .001$, $d = 0.81$, $BF_{10} > 100$. Cueing had a significant effect with moderate evidence on distractor characters, with participants slower to fixate on them during cued trials than uncued trials, $t(44) = 3.26$, $p = .010$, $d = 0.53$, $BF_{10} = 5.35$.

In summary, when frequent threats were cued, participants were faster to attend to them and spent longer looking at them during the trial than on uncued trials. Distractor characters showed the opposite pattern: They were fixated upon more rapidly and for longer during uncued than cued trials.

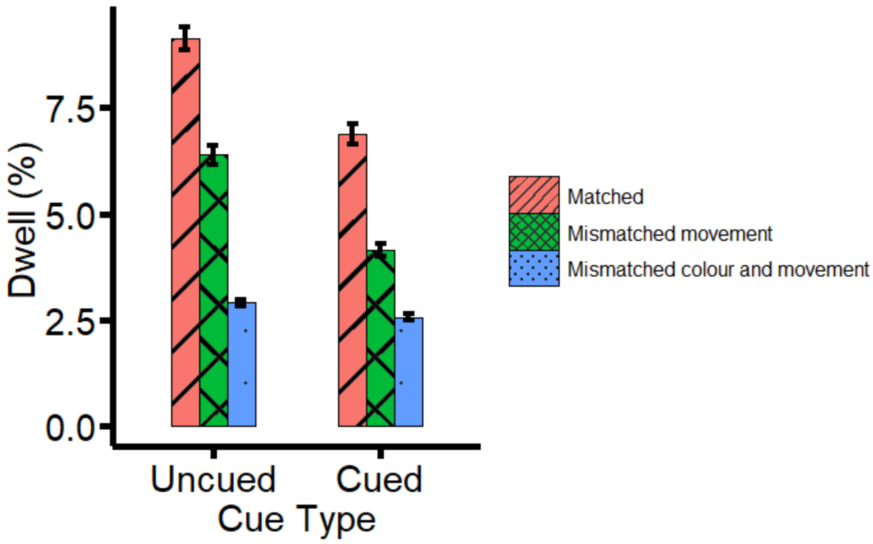
Distractor Type Comparison

I further performed a detailed analysis of distractor types in order to better understand the visual features that participants were using to guide their search during the task. Such an analysis was possible because the subtypes of distractors varied from each other with respect to how closely their shirt colour and movement patterns matched those of the frequent threat characters. Suppose people used these visual features (shirt colour and movement pattern) to perform their search for the threat characters. In that case, we might expect that distractor characters that share more perceptual features with the search target to be gazed at more than characters without these features. Specifically, movement and colour-matched distractors (n = 4) would attract more gaze than other subtypes, such as the mismatched movement (n = 4) and mismatched colour and movement (n = 12) distractors. Moreover, if salient, valid and reliable visual cues overshadowed the more subtle shirt-colour and movement features, then these distinguishing visual cues (shirt colour and movement pattern) should be more useful in the uncued than cued condition

These hypotheses were examined by categorising distractors into three subtypes: colour and movement matched, mismatched movement and mismatched colour and movement. Mean dwell time per distractor subtype is plotted in Figure 12. I restricted gaze measurement to the primary gaze measure of dwell for these secondary, exploratory analyses. I performed a repeated measures ANOVA to evaluate the effect of trial type (uncued/cued) and distractor type (Mismatched colour & movement/Mismatched movement/Matched) on dwell. The main effect of distractor type was significant with extreme evidence, $F(2, 88) = 372.68$, $p < .001$, $\eta^2_G = 0.75$, $BF_{10} > 100$. The interaction between character and trial types was significant (Figure 12), $F(2, 88) = 3.93$, $p = .023$, $\eta^2_G = 0.02$, $BF_{10} > 100$. The main effect of cueing was not considered here as it would repeat outcomes from the frequent threat vs distractor analyses above.

Figure 12

Dwell by Cue Type Per Distractor Category.



Note. Error bars represent standard error.

Instead, the effect of cueing was examined at the stimulus level. Paired samples t-tests were used to compare the effects of cueing on individual distractor types. Results revealed a significant effect with extreme evidence of cueing on the dwell on matched distractors with greater dwell during uncued trials than cued trials, $t(44) = 6.68$, $p < .001$, $d = 1.00$, $BF_{10} > 100$. A significant effect with extreme evidence of cueing was also observed for mismatched movement distractors with greater dwell during uncued trials, $t(44) = 9.43$, $p < .001$, $d = 1.41$, $BF_{10} > 100$. A significant effect with strong evidence was observed on mismatched colour and movement distractors, $t(44) = 3.67$, $p < .001$, $d = .55$, $BF_{10} = 42.70$. These results suggest that regardless of a cue's presence, the closer a distractor matched the frequent threat character in shirt colour or movements, the more participants attended to them. However, the interaction contrast revealed that the magnitude of this gaze bias was stronger in the uncued condition, in which no other visual cues were available to guide the search.

Global Search Behaviour

Exploratory analyses were performed on participants' global search behaviour during the critical trials. Recall that cueing could be predicted to have quite different effects on scanpath: a more biased fixation distribution could lead to a shorter scanpath via each saccade being more direct and efficient, or it could allow people to successfully make larger individual transitions (saccades) between search targets, thereby leading to longer scanpaths. To disambiguate these interpretations, I additionally quantified the number of fixations participants recorded per second and the spatial distribution of their fixations, or search "breadth."

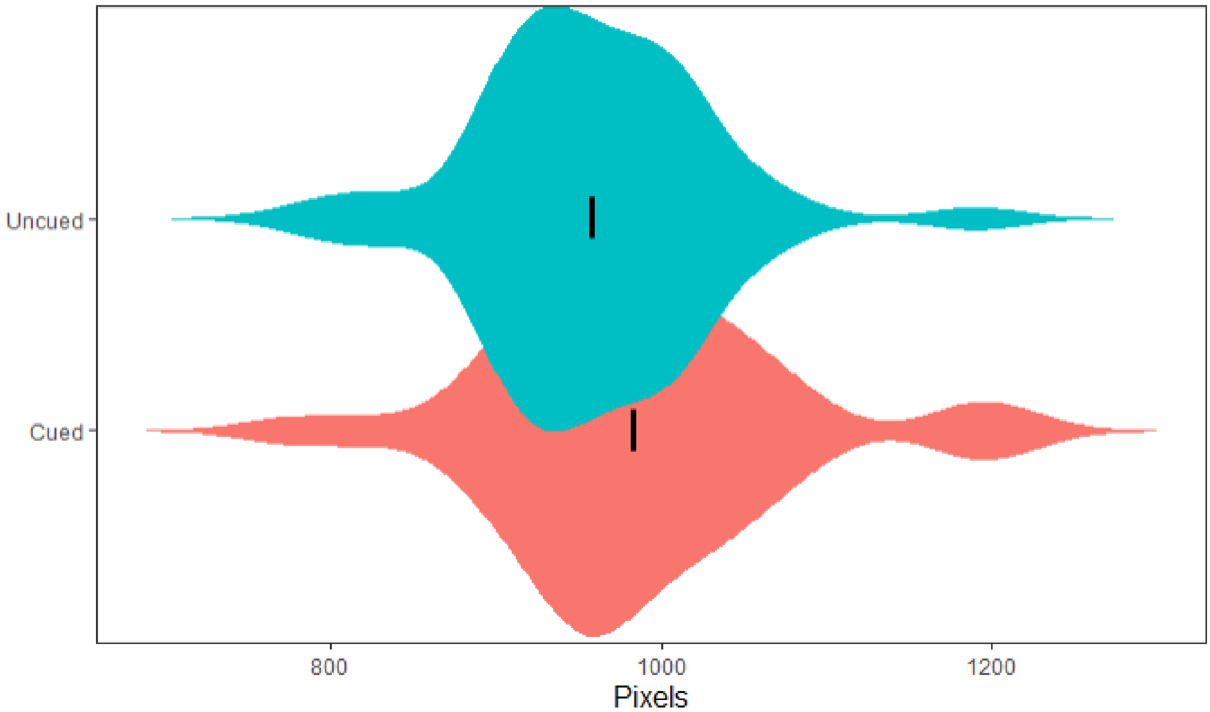
With respect to scanpath, a paired samples t-test showed an effect of cueing during the critical trials with fewer scan pixels per second during cued trials ($M = 1147$, $SD = 180$) compared to uncued trials ($M = 1187$, $SD = 184$), $t(44) = 2.24$, $p = .030$, $d = 0.33$, $BF_{10} = 1.54$. This effect was small, with anecdotal level evidence.

The examination of fixations per second revealed an effect of cueing during the critical trials. There were fewer fixations per second during cued trials (*M* = 2.83, *SD* = 0.70) than uncued trials (*M* = 2.94, *SD* = 0.70), $t(44) = 2.48$, $p = .017$, $d = 0.37$, $BF_{10} = 2.46$. This effect was small, with anecdotal level evidence. In summary, during the cued trials, people's gaze points transitioned across fewer pixels per second and registered fewer fixations than in the uncued condition.

Turning to the question of search breadth, two distinct metrics were obtained. First, the minimal ellipse that covered 95% of gaze points had greater width during cued trials than uncued trials, $t(44) = 3.78$, $p < .001$, $d = 0.56$, $BF_{10} = 57.70$. This effect was moderate, with very strong evidence. The mean and distribution of these width values are summarised in Figure 13.

Figure 13

Violin Plot of Ellipse Width in Pixels by Cue Type During Critical Trials.

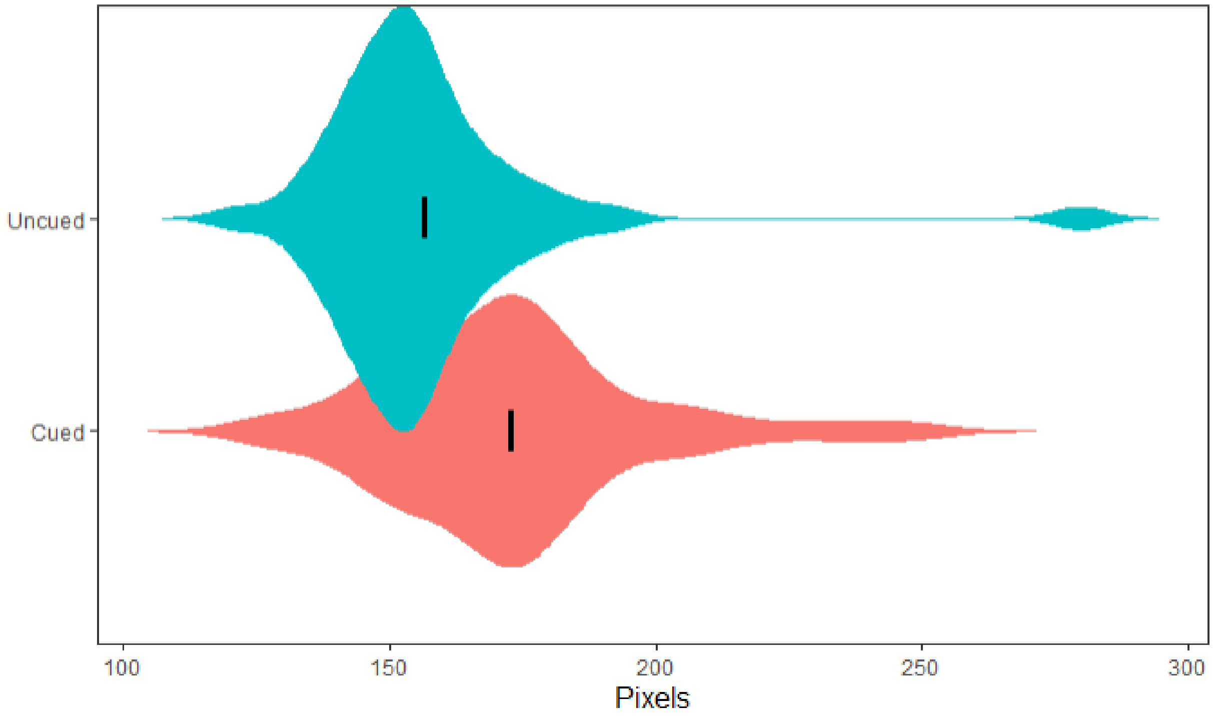


Note. Bars represent the mean ellipse width.

The second metric was more task-specific. This metric examined the mean distance of each gaze point to the nearest friendly soldier (that is, the nearest high-likelihood target region) during the critical trials. The mean distance per cueing condition is summarised in Figure 14. A paired samples t-test revealed an effect of cueing with the distance to the nearest friendly soldier greater during cued trials than uncued trials, $t(44) = 5.42$, $p < .001$, $d = 0.81$, $BF_{10} > 100$. This effect was large with extreme evidence.

Figure 14

Violin Plot of Distance in Pixels of Fixations From Nearest Friendly Soldier Character by Trial Type During Critical Trials.



Note. Bars show the mean distance in pixels.

Both search breadth analyses show that participants spread their gaze wider during cued trials than uncued ones. The distance to the nearest friendly measure (Figure 14) highlights this difference more clearly than the ellipse width measure (Figure 13), likely because the former measure is designed specifically for the present task, whereas the ellipse measure is more general. Combining outcomes from the scanpath, fixations and search breadth measures shows that participants are spreading their gaze broader whilst having fewer fixations and moving less distance between each fixation during cued trials. There was no clear prediction for these outcomes, however the results align with one possibility discussed in the Global Gaze Behaviour section above. That if cued characters attract attention and those characters frequently onset at locations distant from one another then gaze is likely to oscillate between those onset locations. Reducing gaze towards uncued characters, increasing gaze towards cued characters, and increasing search breadth.

Full-Attention Trials

The full-attention control trials provided an opportunity to investigate if the rare threat character was plainly visible when participants where unburdened by the demanding frequent threat search task and where expectancy should be relatively high due to having completed the critical trials. Full-attention trial rare threat character data was grouped into a 2 (Hits, Misses) x 2 (uncued, cued) table (Table 3). As can be seen in the table, misses were relatively rare during the full-attention control (4% overall). A G-test showed that the proportions of hits and misses between cued and uncued full-attention trials were not significantly different $G(1) = 2.45$, $p = .117$, $V = .09$, $BF_{10} = 0.36$.

Table 3

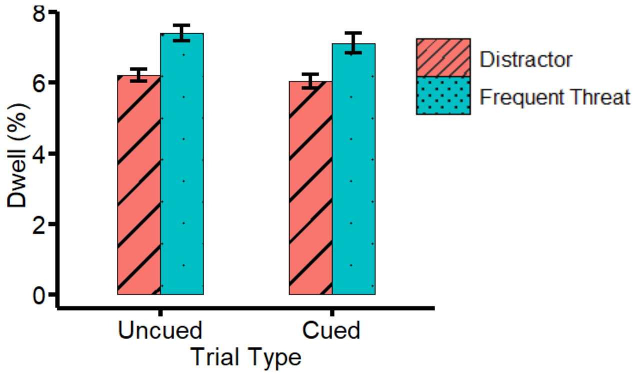
Contingency Table Showing Rare Threat Character Hits and Misses for Left and Right Character Onset Locations for Full-attention Trials.

	Uncued	Cued	Total
Hits	135	130	265
Misses	3	8	11
Total	138	138	276

A repeated measures ANOVA was performed to evaluate the effect of trial type (uncued/cued) and character type (distractor/frequent threat) on dwell (Figure 15) during the full-attention trials. The main effect for character type was significant, with extreme evidence that participants dwelled longer on frequent threats than distractors, $F(1, 43) = 26.77$, $p < .001$, $\eta^2_G = 0.13$, $BF_{10} > 100$. The main effect of cue type on dwell was not significant, $F(1, 43) = 1.04$, $p = .314$, $\eta^2_G = 0.00$, $BF_{10} = 0.27$. The interaction between character and cue types on dwell was not significant, $F(1, 43) = 0.12$, $p = .731$, $\eta^2_G = 0.00$, $BF_{10} > 100$.

Figure 15

Dwell by Cue Type for Frequent Threat and All Distractor Characters During Full-Attention Trials.



Note. Distractors include all distractor types. Character type dwell as a percentage of trial length. Error bars represent standard error.

These results show that the rare threat characters are plainly visible during the full-attention trial when not burdened by the demanding frequent threat search task. Furthermore, the effects of cueing on both the frequent threat and distractor characters observed during critical trials disappeared during the full-attention trials.

Rare Threat Onset Location

Anecdotal feedback from experimenters indicated that the gun carried by the rare threats may have been more visible for characters moving from left-to-right versus right-to-left on the screen. As seen in Figure 16B, when moving from right-to-left, the character's body obscures the gun, an artifact of the character's body rotation and participant line of sight, as the gun was always carried in a fixed, right-handed position. This fixed gun position is more clearly seen in the left-to-right travel shown in Figure 16A.

Figure 16

Left (Panel A) Screen Onset vs Right (Panel B) Screen Onset Comparison of Rare Threat Character.

A B

Note. Image contrast adjusted for publication purposes.

To examine this question we conducted exploratory analysis by splitting the appearances of the rare threat character into left or right onset paths for all trials they appeared in. Each participant had twelve appearances of the rare threat character across the critical and full-attention trials, with the onset side randomly selected by the experiment for each appearance. We then used paired samples t-tests to examine onset counts, response times, and miss rate performance for rare threat left vs right onset appearances (Table 4). The response time and miss rate performance measures observed significant differences due to screen-edge onset. Fortunately, there was no significant difference between the number of left and right onsets per participant.

Table 4

Paired Samples T-Test Results Comparing Left and Right Screen Onset Measures for Rare Threat Characters.

Measure	*Mean (SD)*		*t (df)*	*p*	Cohen's *d*
	Left	**Right**			
Onset count	6.22 (2.00)	5.78 (2.00)	0.74 (45)	.465	0.11
Response times (sec)	2.90 (1.04)	5.48 (2.44)	6.40 (44)	<.001	0.95
Miss rate (%)	11.46 (22.30)	27.97 (25.46)	5.24 (45)	<.001	0.77

Note. Data collapsed over all trials with rare threat appearances, i.e., trials 3-6—maximum of 12 appearances per participant.

The full-attention control trials provided an opportunity to investigate if left vs right onset appearances remained a concern when participants were not required to perform the demanding frequent threat search task and where expectancy should be high due to having completed the critical trials. Full-attention trial data were grouped for the rare threat character into a 2 (Hits, Misses) x 2 (Left onset, Right onset) table (Table 5). A G-test showed that the proportion of misses between rare threat character left and right onset locations during full-attention trials were not significantly different, $G(1) = 1.57$, $p = .210$, $V = .07$, $BF_{10} = 0.23$. When expectancy was high, experience with the task was high, and task demands were low, no left/right onset differences were observed. However, differences due to screen onset did emerge under the more challenging initial conditions in which the critical trials occurred.

Table 5

Contingency Table Showing Rare Threat Character Hits and Misses for Left and Right Character Onset Locations for Full-attention Trials.

	Left	Right	Total
Hits	143	122	265
Misses	8	3	11
Total	151	125	276

Trust and Accuracy Perceptions

Trust (*M* = 4.13, *SD* = 0.82) and perceived accuracy (*M* = 86.30, *SD* = 15.13) of the cues were strongly correlated ($r = 0.75$, $p < .001$) and generally indicated high levels of trust. Only 6% of participants rated trust in the cues below three on the scale (max. five), and only 4% indicated the perceived accuracy of cues as less than 70%. This is likely due to the cue being 100% valid for frequent threat characters on all trials in which it is presented.

Inattentional Blindness Awareness

30% of participants reported knowing about inattentional blindness before this experiment, and 65% reported knowing about the Gorilla experiment (Simons & Chabris, 1999). No participants indicated associating this experiment with inattentional blindness or the Gorilla experiment.

Discussion

Prior research has indicated that AR visual cues may offer visual search performance benefits for cued threats but at a cost to uncued threats. This experiment aimed to investigate this potential cost/benefit trade-off within the context of a video game simulation of a dismounted soldier overwatch scenario. Participants were tasked with searching for any threat, with a threat defined as anyone carrying a weapon. Unknown to the participants, there were two types of threats, frequent and rare. Participants performed the overwatch aided by 100% reliable AR visual cues on frequent threats in half of the trials, with no cues provided in the remaining trials. Rare threats were not cued in any trials. The design was based on traditional inattentional blindness paradigms (Mack & Rock, 1998).

The cueing implementation was effective. Cueing led to predicted improvements in search performance and increased attention towards the cued frequent threats. The cueing of the frequent threat characters resulted in a decrease of attention towards distractors and an increase in attention towards the cued frequent threats. The improved behavioural responses for frequent threat characters during cued trials confirm that the cues performed as designed, i.e. improved search for cued stimuli. This is an important check to ensure the experiment's cues functioned as expected even though such effects are well documented (see Posner, 1980 for a review).

The primary hypothesis was that providing cues during an applied visual search task would lead to performance costs for uncued threats due to attentional tunnelling to cued threats. But this hypothesis was not supported as visual search performance for the *uncued rare* threats was not significantly different during critical trials when *frequent* threats were cued. This conclusion held irrespective of whether we analysed all appearances of the rare threat character or just the initial appearances. This conclusion also held regardless of if we considered the primary inattentional blindness measure of rare threat miss rate, the response time or the gaze measures of dwell and time to first fixation during critical trials. Cueing showed no effect on behavioural or gaze responses.

Despite not seeing statistically significant outcomes, a consistent pattern of non-significant, small numerical effects suggested that there may be an attentional cost to uncued threats during cued trials, which the present experiment could not detect. For example, when the detection analysis is limited to the first presentation of the rare threat, as is typical in traditional inattentional blindness studies, the proportion of those classified as non-noticers (i.e., inattentionally blind) rose from 55% in the uncued trial to 71% in the cued trial. Although this is the typical method to measure inattentional blindness, it presents challenges to inferential analyses. First, it relies on a single measurement and is particularly susceptible to measurement error. Second, we observed a reliable difference in the detectability of left- and right-onsetting characters, and a single measurement is highly sensitive to the noise this difference would induce. So, although this difference was not significant in the present experiment, the numerical magnitude (16%) of the putative effect of cueing is comparable to the reported magnitude of other successful manipulations on inattentional blindness (see Mack & Rock, 1998) and warrants further investigation in variants of this task that are designed to minimize statistical noise.

Cued targets likely induced attentional tunnelling. This was evident in faster reaction times to frequent threats during critical trials, and a larger gaze bias towards these stimuli when cueing was present. Moreover, the impact of non-cueing perceptual features (shirt colour or movement pattern) on gaze to distractors was weakened when cues were presented, suggesting people relied predominantly on cues when they were available.

Faster times to first fixation for cued frequent threats during critical trials also suggests participants may have engaged in preattentive search processes towards cued stimuli; the cued stimuli “popped out”. This combination of results implies that attentional tunnelling occurred, with cued characters prioritised for earlier attention than uncued characters and those same characters receiving more attention.

Crucially, participants were tasked with *searching for weapons.* They were not instructed to search for salient cues, particular weapons (knives), or perceptual features associated with their targets (red shirts or character motion patterns). The instruction was clear and direct, and comprehension was checked before commencing the task. So, whilst this increased attention towards the cued threats improved search performance for those threats, the failure to appropriately disengage from the cues led to a failure to allocate attentional resources to the actual task described to participants (the detection of weapons). This could be understood as a missed opportunity. When cues were present, participants performed the search for frequent threats very well: rapid reaction times, low miss rate and highly preferential gaze allocation. One interpretation of this finding is that this facilitated performance ought to have left them with more residual capacity to detect unexpected or novel events. Yet, if anything, the reverse pattern was observed.

Within the present experiment, participants are provided with a visually salient and task-relevant cue. Therefore, it is feasible that participants may have activated a feature-based attentional set based on the colour or size of the visual cue but also may have prioritised the cue due to its task relevance. Leaving the question: was the attentional tunnelling observed due to the cue’s visual salience, task relevance, or both?

Visual search literature shows that people use a feature-based attentional set to simplify search (e.g., Simons & Chabris, 1999; Treisman & Gelade, 1980; Wolfe, 2007). Such a set is created by identifying features that differentiate target stimuli from distractor stimuli, including features such as colour, size, shape, luminance, or combinations. The general finding in this literature is that applying an attentional set leads to improved search outcomes for stimuli whose features more closely match the selected attentional set and poorer outcomes for stimuli less closely matched (Most et al., 2005). Researchers have also shown that participants can prioritise attention to stimuli based on task relevance or value (Le Pelley et al., 2015; Watson et al., 2020).

The highlighting cues were both salient and subjectively relevant. The salience of the bright red overlays is immediately evident; see Figure 1 for an example. Trust ratings were used to assess the cues' perceived task relevance. The cues were highly trusted: across all participants and all trials, the cues were rated as 86% accurate. Within decision aid research, aids perceived as more than 70% accurate are far more likely to be used (Wickens & Dixon, 2007). The present highlights exceeded that threshold.

The results from the full-attention trials were reviewed to determine if participants prioritised cue visual salience or task relevance. One reason for the full-attention trials inclusion in the experiment was to see how attention towards frequent threats changed when the cues were no longer task-relevant but remained visually salient.

The most important finding was that the effect of cueing on frequent threat gaze measures observed in the critical trials (Figure 11) disappeared in the full-attention trials (Figure 15). A supporting result is that the dwell on frequent threat characters was ~250-300% larger than dwell on distractor characters during the critical trials (Figure 11). However, within the full-attention trials, the difference between the frequent threat and distractor character dwell times reduced to less than 20% (Figure 15). This sizable reduction suggests that frequent threat characters were attended to during full-attention trials more like distractors than targets. Participants were not directing attention towards frequent threats to the same degree as in critical trials. As the only frequent threat attribute changed between these trials was that they were no longer task-relevant, this indicates that task relevance was the primary motivation for attending to frequent threat characters more quickly and for longer in critical trials. Objective stimulus features (such as size or brightness) were less important.

Attentional Sets

Despite task relevance being the primary motivation for the increased attention towards frequent threat characters, there is also some evidence that an attentional set for red shirts was activated. The pattern of results seen in Figure 12 shows that red distractors (Matched and Non-movement matched distractors) received more attention than non-red stimuli (Mismatched distractors) in both the cued and uncued critical trials. These red-coloured distractors were dwelled on for longer and were faster to be fixated upon than non-red distractors. Whilst the provision of cueing for frequent threat characters reduced attention towards distractor characters, when attention was directed towards distractors, red distractors were prioritised over non-red distractors. The matching red shirts of some distractors and the frequent threat characters were a deliberate design decision to reduce the visual differences between these character types. This design did not function as hoped as participants likely used the red shirts to guide search. Therefore, in future studies, shirt colours have been reconsidered.

Distractor movement matches to frequent threat characters was also considered a potential attentional set. As seen in Figure 12, grouping distractor characters based on similarity in colour and movement to the frequent threat characters suggests an interesting role for motion in any feature-matching process. The figure shows that regardless of cueing, distractors that match both colour and movement of the frequent threat characters receive more attention than characters who only match colour, who receive more than characters with no match. However, screen location challenged interpretations from these movement match findings. Characters that share common trajectories are more likely to be co-located in space, and thus any given fixation that falls in one character's ROI is likely to also fall in the ROI of another converging character. This was a potential problem because 36% of all fixations were assigned to two or more characters due to overlapping character ROIs at the fixation point.

However, disambiguating the role of common location on a background of common movement trajectories is complex, and the present experiment was not primarily designed to address this question and so lacks relevant experimental controls.

Rare Threat Onset Location

While cueing did not change detection rates for the rare threat, the side of the screen that the character entered from did influence detection. As shown in Figure 16, the right-handedness of the rare threat character rendered the rifle more visible when he entered from the left. Inferential analyses of detection rates and speed confirmed this observation. It is presumed that this laterality effect is due to the visual appearance of the rifle, but it is also possible that the mere positioning of this character might have influenced detection. Specifically, some evidence shows that response times are faster for stimuli appearing in the left vs right visual field. However, these findings show differences in the millisecond range (e.g., Woods et al., 2015), and thus 2-3 orders of magnitude smaller than that observed here. In all subsequent experiments, the rare threat character alters his handedness to match his side of onset, thereby equating the visibility of the rifle across right and left entrances. To anticipate, the effect of laterality is subsequently non-significant in all subsequent experiments. Consequently, we will not return to this issue.

Global Search Behaviour

Global search behaviour was an exploratory area within this first experiment. Predicting how participants may interact with the display was difficult as neither consistent scanpaths nor an optimal scanpath was expected due to the complexity of the scene (c.f. Frame et al., 2019), and these attributes are often required for the interpretation of derived scanpath metrics (Gilchriest, 2011). More concretely, experiments that use scanpath metrics typically have relatively few stimuli and relatively little movement so as to enable precise interpretations. The present experiment did not possess these features because it was modelled on an applied search task, in which visual complexity and movement is commonplace. Nevertheless, we sought to quantify and interpret search behaviour using scanpath under these atypical conditions.

Global search behaviour analyses provide important insights into *how* the decision aid (highlighting, in this instance) was used. Evidence within all four global search measures suggests that cueing altered how participants searched the screen.

When cues were present in critical trials, participants employed a more efficient search strategy with shorter scanpaths and longer fixations. This efficiency was attained by more direct saccades from one character to the next combined with fixating on those characters longer. Interestingly, the shorter scanpaths in the cued condition occurred despite a broader search distribution. Superficially, one might expect broader gaze distribution to require longer scanpaths. Yet when viewed collectively, our metrics reveal that cueing allowed more direct gaze transition between target characters (than uncued conditions), allowing the gaze point to traverse to more broadly distributed regions of the screen without people moving their eyes more. Put another way, gaze paths were less direct (more “zig-zag”) in uncued conditions, as people fixated on distractor characters more on their way to potential target characters, and this resulted in people being less able to visit distal regions of the screen in their search for targets. Here, cuing did not just facilitate target detection but allowed participants to foveate a larger proportion of the screen area across the course of a cued trial.

However, there is an important caveat to this performance improvement. Within the context of the present experiment, a more biased (or “efficient”) search for cued characters is likely a suboptimal strategy when viewed holistically. Indeed, it is analogous to the attentional tunnelling reported in operational search literature (e.g., Wickens & Alexander, 2009). An efficient strategy for search tasks with known targets involves short scanpaths with fewer, more direct saccades to targets (Duchowski, 2017). Saccading to and fixating on distractors is an inefficient use of resources in a world in which none of those uncued distractors are meaningful. However, that was not the case in the present task (nor in many real-world search tasks), and this resulted in less opportunity to identify those characters as carrying a weapon. Rephrased from the perspective of operational search, participants tunnelled their attention towards cued targets, improving performance for those cued targets, but crucially it also increased the potential for blindness to uncued targets.

Within this experiment, I used two measures of search breadth: ellipse and distance to nearest friendly. Both measures indicated similar outcomes on search breadth, and in the interests of parsimony, I determined to use the measure optimised for the present search task: distance from nearest friendly soldier.

Inattentional Blindness Awareness

The fame of the Simons and Chabris (1999) Gorilla experiment raised concerns that participants may not have been naïve to inattentional blindness and may have deduced the present experiment's focus. Post-experiment questioning of participants indicated that it was a valid concern, as 65% of participants were aware of the Gorilla experiment, and 30% aware of inattentional blindness. However, I concluded that the participant pool was sufficiently naïve because no participants indicated that they related this experiment to the Gorilla experiment or inattentional blindness until after answering these questions. Further, analyses of the first presentation of the rare threat character showed that most participants (approximately 60%) failed to detect the rare threat.

The degree of naivety reported post-experiment was important for the experiment's validity but also surprising. Within common inattentional blindness studies, participants have only one or two appearances of the unexpected stimuli, therefore less opportunity to associate the experiment with inattentional blindness research. However, within our experiment, they experience 12 appearances of the unexpected stimulus. Yet they still did not make the connection that this is an inattentional blindness experiment, increasing my confidence that the patterns of responding observed were not a result of prior knowledge of the experiment's intent.

Conclusion

The addition of Augmented Reality visual cues improved performance for cued targets at no apparent cost to uncued targets, i.e., inattentional blindness did not increase. However, secondary analyses revealed strong evidence that cueing changed the nature of search. People traversed from one cued object to the next with little time spent fixating on uncued objects, thereby increasing the likelihood of missing uncued threats. While this was not evident in detection rates for rare threats, analysis of detection rates for the initial onset of the rare threat showed a numerically (but not significantly) lower detection rate in cued than uncued conditions. Moreover, the slight imbalance in visibility of left- and right-entering rare threats, and the differential rates of distraction afforded by the varied distractor types may have contributed experimental noise that reduced sensitivity to any effect of cueing on inattentional blindness. A second experiment addresses these design issues, increased sample size, controlled for the salience of the cueing stimuli themselves and introduced a neural measure of covert-attention.

Chapter 3: Experiment 2

Introduction

Experiment 1 simulated a military overwatch scenario. The primary research question concerned quantifying both the benefits and costs of introducing valid, reliable, onscreen cues for characters identified as a potential threat. Overall, cueing assisted detection and response speed for highly frequent, cued threats. This benefit did not seem to incur any costs in the speed or accuracy of detecting rare, unexpected, and uncued threats. However, there are reasons to believe that the cost of cueing may have been obscured by small design choices within the video game simulation. Changes such as character shirt colours have been implemented for the present experiment to address these design concerns. The rationale for larger changes is discussed sequentially below and in the method section for minor changes.

The largest change from Experiment 1 lies in the broad design, with Experiment 2 adopting a between-subjects design with a single appearance of the rare threat and introducing a new control group. The use of a between-subjects design narrowed measurement to the first occasion in which the rare threat was shown. The largest determinant of whether a rare threat would be detected or not in Experiment 1 was whether a rare threat had previously been detected; people tended not to miss rare threats after they had successfully detected one rare threat. This aligns with the repeatedly demonstrated role of expectancy in shaping when inattentional blindness occurs (Mack & Rock, 1998). For this reason, Experiment 2 follows the classic inattentional blindness protocol more closely, with a single critical trial that is then repeated as a single full-attention trial with the requirement to perform the primary search task removed. Quantification of inattentional blindness now focuses on the first time the rare threat occurs, when its expectation is unambiguously low.

A second cueing condition was introduced to resolve an inherent confound between the cued and uncued conditions used in Experiment 1 (and elsewhere, Yeh & Wickens, 2001). The comparison between cued and uncued conditions confounds the presence of visual cues with their utility. That is, any effect of cueing might have been due to the cues' high validity or simply due to the addition of bright, salient objects on the screen. Such visual objects might increase engagement (and thus improve arousal and performance) or induce visual clutter (thus hampering search performance). To directly test these hypotheses, the present experiment extends the cued and uncued conditions with a third condition: uninformative cues. The new *uninformative* cue condition cues as many characters as the cued condition (hereafter referred to as the *informative cue* condition). However, in the *uninformative* condition, the cues were perfectly non-contingent (Rescorla, 1967), with only 3 of 9 cued characters being frequent threat characters, the remainder being distractors. That is to say, the presence of cues provided no information above the baseline rate of threat in the scene. Note that this control differs from a purely invalid cueing protocol, in which cues are exclusively placed over distractor stimuli. Such a protocol is paradoxically informative: once understood, it tells people where *not* to look. A zero-contingency or uninformative protocol does not possess this property.

With the introduction of uninformative cues, a further question arises concerning which aspect of the cues provided is expected to impact inattentional blindness, salience or task relevance. Suppose salience is the critical determinant of attentional capture in this task. In that case, inattentional blindness should be larger whenever cues are present: informative or uninformative. The alternate determinant of attention capture is whether cueing can funnel attention towards task-relevant cues by their continued informativeness and validity. In that case, only informative cueing should increase errors of inattentional blindness.

The same influences are predicted to alter behavioural outcomes for frequent threat characters. Behavioural outcomes towards frequent threats will improve in the presence of informative cues regardless of if salience or task relevance is the key determinant of attention to the cues. As either account predicts that attention will be captured by the 100% reliable cues for frequent threat characters in this condition. In contrast, the presence of uninformative cues was predicted to worsen if salience is the determinant of attention capture, as attention will be drawn to invalidly cued distractor characters. Whilst under a task-relevance account of attention capture by cues, uninformative cues should fail to attract attention (as they lack task-relevance), and frequent threat responses should be similar to the uncued condition.

Finally, the present experiment altered the perceptual features of onscreen cues to allow the measurement of covert attention. Specifically, a constant 15Hz sinusoidal flicker was induced within the red overlay cues to elicit an SSVEP. SSVEPs are measured using electroencephalography and are commonly used to measure covert or feature-based attention (see Norcia et al., 2015 for a review).

Overt vs Covert Attention

Understanding overt and covert attention during visual search is important because they refer to different attentional processes, which may be important. Overt attention in a visual search task can be measured via eye gaze analysis (e.g., Richards et al., 2012), only measuring what is visible in the high-definition foveal visual field. Answering questions such as, what are participants looking at, for how long, and how often? However, objects may appear anywhere in the visual field, including peripheral vision. Objects have a lower visual definition the more distant they are from foveal vision, but these objects may still garner attention, referred to as covert attention (Posner, 1980).

Covert attention is considered the precursor to overt attention. More specifically, for visual attention, it is suggested that an object can draw covert attention from anywhere in the visual field, including peripheral vision. Potentially triggering a saccade to the stimuli which is then overtly attended to (Posner, 1980). Within the current research, the high visual salience of the AR cues was expected to attract covert attention regardless of location in the visual field, subsequently triggering a saccade to the cue's location. Evidence supporting this expectation was observed in Experiment 1 when eye-tracking indicated participants more rapidly attended to cued objects after screen onset than uncued objects.

Steady State Visually Evoked Potential

Covert attention is difficult to quantify (Duchowski, 2017), but one promising method uses the SSVEP technique (e.g., Walter et al., 2012). SSVEP uses EEG to measure entrainment to an external flickering stimulus (Norcia et al., 2015). When a flickering stimulus is viewed, occipital scalp sites detect neurocortical oscillatory activity aligned with the phase and frequency of that flickering stimulus. Furthermore, this obligatory neural response – the SSVEP – is modulated by attention (Walter et al., 2012). When the flickering stimulus is attended to, the magnitude (power or amplitude) of the SSVEP response increases relative to when it is ignored. Moreover, this modulation occurs for stimuli presented extra-foveally, for stimuli that the person is not currently gazing towards (Walter et al., 2012). In this manner, it allows attention measurement beyond the current gaze point and thus complements overt measures of attention, like eye tracking (see Griffiths et al., 2021).

Notably, SSVEP measurement has other attributes suitable for this particular task. First, unlike gaze-based measures, SSVEP allows attention to be measured simultaneously to several cues present across the visual field as long as they are all flickering simultaneously (Norcia et al., 2015). This allows moment-by-moment measurement of the degree to which participants monitor all eight visually spaced cues around the scene, distinct from the gaze point, which is either direct at zero or more cues at any given moment. Second, and perhaps most importantly, SSVEP measures are unaffected by spatial overlap (Norcia et al., 2015). In Experiment 1, participants commonly (36%) recorded fixations within multiple regions of interest (ROI). This occurred because characters in the scene moved in front of each other, particularly in the central regions of the display. Gaze then becomes difficult to interpret:

Were people attending to cued character A, or the cue hovering over character A, or towards the spatially collocated uncued character B? By contrast, SSVEP is only affected by the flickering stimulus, so any change in SSVEP magnitude can only be induced by a shift in attention (overt or covert) to the flickering cueing stimulus, and not by the characters these cues overlaid.

The SSVEP signal magnitude is measured as power in the frequency spectra at the bins nearest the target frequency (i.e., the rate at which relevant stimuli were flashing at). All steps involved in calculating this metric are reported in the Methods, but in essence this power metric is normalized to the baseline level of power in that frequency band when no flickering stimulus is present, and so is best thought of a stimulus-locked increase in power.

In other respects, the present experiment resembled Experiment 1, and addressed some of the same hypotheses.

Hypotheses

- Rare threat detection performance was predicted to be lower in the informative cue condition compared to the uninformative and no-cue conditions, with higher miss rates and response times.
- Frequent threat search performance was predicted to be higher in the informative cue condition compared to the uninformative and no-cue conditions, with lower miss rates and faster response times.
- Overt attention (dwell) towards rare threat characters was predicted to be higher in the no-cue condition than the uninformative cue condition, which was further predicted to be higher than the informative cue condition.
- Overt attention (dwell) towards frequent threat characters was predicted to be higher in the informative cue condition compared to the uninformative cue condition, which was further predicted to be higher than the no-cue condition.

- SSVEP power was predicted to be higher in the informative cue condition than the uninformative cue condition.

Method

Participants

Ninety-four participants from the Flinders University First Year Psychology undergraduate participant pool (69 Female, 25 Male; mean age 22.27 years, *SD* = 5.93) participated in exchange for course credit. All participants gave informed consent and during pre-screening reported normal or corrected-to-normal vision and normal colour vision. This research complied with the tenets of the Declaration of Helsinki and was approved by the Social and Behavioural Research Ethics Committee at Flinders University (Project ID: 2569).

COVID-19 related restrictions in place throughout the experimental period largely dictated access to participants and therefore final participant numbers. A priori power analysis conducted using G*Power3.1 (Faul et al., 2009) for a χ^2 test, $\omega = 0.30$, $df = 1$, Power = 0.80 and *alpha* = .05 indicated a sample size of 88. The effect size was estimated based on the results from the first critical trial rare threat noticing rates from Experiment 1, which was expected to most closely match the present experiment design. This effect size was $d = 0.32$, which is interpreted as a small to medium effect (Cohen, 1988). A medium effect size was selected for the present analysis due to design changes that were expected to increase noticing rates. In the absence of a specific G-test power analysis the chi-squared power analysis was considered a suitable analogue as the G-test and chi-squared tests are functionally similar (Agresti, 2019).

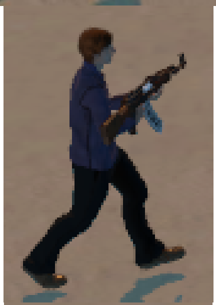
Stimulus and Materials

The experiment was adapted from Experiment 1, with design alterations documented in the remainder of this method section. Shirt colour assignments in Experiment 1 resulted in poor distractors due to a mismatch with frequent threat characters. Therefore, two character-specific changes were made. The character counts for this experiment were increased from 33 to 35 characters, enabling equal distribution of shirt colours within frequent threat and distractor characters (Table 6). Character shirt colours were made more homogeneous with new colour and character type combinations summarised in Table 6.

Noticing rates during critical trials in Experiment 1 were quite high (91%), reducing statistical sensitivity. Three changes were therefore introduced to increase the difficulty of the primary search task. Firstly, the previously discussed increased homogeneity of shirt colours. Secondly, the rate of frequent threat attacks on friendly soldier characters was increased from an average of one attack every ~4.2 seconds to one per ~2.4 seconds, thereby increasing the cognitive load induced by the primary task. Finally, the character paths were altered from Experiment 1 such that characters no longer left the scene. Rather than traversing a waypoint network, then exiting the screen and pausing before re-entering, characters now continually moved towards waypoints on the screen throughout the whole trial. This meant that the number of distractors present at any moment was constant.

Table 6

Character Types and Attributes.

	Qty	Character Type	Gender	Shirt Colour(s)
	9	Frequent Threat	Male	3 x Brown 3 x Blue 3 x Violet
	1	Rare Threat	Male	Brown, Blue, or Violet chosen randomly before each screen onset
	4	Friendly Soldier	Male	Grey Camouflage
	21	Distractors	6 x Female 15 x Male	7 x Brown 7 x Blue 7 x Violet

In the preceding experiment, the non-deterministic pathways traversed by each character meant that there were periods in which many distractors were present and periods when few were. Because the present experiment used a between-subjects design in which the rare threat appeared only once on the critical trial, it was important to make sure this presentation did not appear at a time of relative difficulty or ease for one participant but not another, thereby introducing statistical noise into the measurement. For this reason, characters no longer left the screen, resulting in a constant level of difficulty across each trial.

Cue design remained similar to Experiment 1 except for cues being frequency tagged at 15Hz (i.e., flashing at a rate of 15Hz). A BioSemi ActiveTwo Electroencephalography (EEG) system collected data to enable the SSVEP analysis, see Measures section for more detail.

Design

The experiment was a between-subjects design with participants allocated to one of three conditions no-cue (*n* = 32), uninformative cue (*n* = 31) or, informative cue (*n* = 31). The informative cue and no-cue conditions matched the conditions from Experiment 1. All participants completed three trial-types that occurred in the following order: training, critical, full-attention.

The new uninformative cue condition cued nine characters, the same number of characters as the informative cue condition but with a mix of correctly and incorrectly cued characters. It incorrectly cued three female and three male distractors whilst correctly cueing three frequent threats. In summary, within the uninformative cue condition, only one-third of the cued characters were threats, the remainder being distractors.

Procedure

Commencement and instructions to participants remained similar to Experiment 1. Except participants only received cue instructions at the start of the experiment as all participants were assigned to a specific cue type condition, and the cues they experienced were the same during all trials. No further instructions were given between trials. See Table 7 for the pertinent instructions per condition.

Table 7

Specific Instructions by Condition.

Cue Condition	Instructions
No-cue	You are in the no-cue condition. No target detection aids will be present and you will have to identify threats on your own.
Uninformative	You are in the uninformative cue condition. Red cues will appear over some characters who will approach a soldier, pull out a knife and attack. But cues may also appear over characters who are not a threat. Therefore, you must try and ignore the cues.
Informative	You are in the informative cue condition. Red cues will appear over men who will approach a soldier, pull out a knife and attack.

Training Trials

The training trials remained similar to Experiment 1 except that the between-subjects design meant that participants only experienced cueing based on the condition they had been assigned, e.g., those in the informative cue condition received informative cues during all trials.

Unlike Experiment 1, in this experiment, frequent threats did not exit the scene after completing their attacking animation. Instead, they were immediately allocated a new random waypoint-network to traverse before approaching and attacking a friendly soldier again. Approximately 37 attacks on friendly soldiers occurred per training trial (M = 37.22, SD = 10.02). Although frequent threats predominantly remained on-screen during the trial, some waypoints in the Unity 3D world-space existed outside the screen coordinates. Characters could temporarily (<2 seconds) exit the screen when assigned an off-screen waypoint until they were assigned either an on-screen waypoint or to approach a friendly soldier.

Distractor characters behaved in the same way as Experiment 1 with one exception. Similar to frequent threat characters distractors remained on-screen.

Trials lasted approximately 90 seconds (M = 93.91 sec, SD = 23.12 sec), ending when participants had successfully clicked on between 21 and 29 frequent threats. The exact number was randomly selected before the trial commenced. The participants hit and miss totals were displayed on-screen after the trial finished. A rest period was then offered. This trial format was repeated two more times, starting at the eye-tracker calibration stage, for a total of three training trials. Unlike Experiment 1 no measures were taken between trials.

Critical Trial

The critical trial occurred immediately after the third training trial. The participant received no indication this trial was different to the previous three trials. The only difference between this trial and the training trials was that the rare threat character appeared randomly from the left or right of the screen approximately 45 seconds after the trial started. The character entered carrying a large rifle. After onset, the rare threat traversed to the middle of the scene, stopped, faced the participant for two seconds and then proceeded to exit the scene opposite their onset location. The rifle disappeared if the participant clicked on the character, and then the character continued on their path unarmed, now appearing similar to a distractor character. They were visible for approximately 17 seconds (*M* = 17.62 sec, *SD* = 1.35 sec). Trials ended when the rare threat exited the screen resulting in random trial lengths of approximately 81 seconds (*M* = 81.53 sec, *SD* = 0.90 sec). To address weapon visibility concerns during right scene onsets of the rare threat character, the rare threat character now swaps to a left-handed weapon position during right scene onsets and pauses in the middle of the scene, turns towards the participant and then continues on their assigned path. Unlike Experiment 1, where rare threat characters made multiple appearances per trial, the rare threat character only appears once during the critical trial

Full-Attention Trial

The full-attention trial was similar to the critical trial, except that participants were instructed not to search for and click on the frequent threats. The instructions were that the system would automatically click any frequent threats attacking the friendly soldiers. Experimenters also reminded participants to continue looking for any threats. Each trial ended when the rare threat exited the screen resulting in random trial lengths of approximately 80 seconds (*M* = 80.00 sec, *SD* = 1.06sec).

Post Experiment

After the full-attention trial, participants completed the same Naivety, Video Game Experience, A-SWAT and general feedback forms as documented in the Measures section below. Experimenters then provided a verbal and written debrief explaining the purpose of the experiment.

Measures

Behavioural Measures & Character Tracking

Behavioural measures and character movements were recorded as per Experiment 1, except the sample rate for character screen locations doubled to ~120Hz, matching the screen refresh rate.

Gaze Measures

Gaze measures were recorded and processed, equivalent to those from Experiment 1. However, as the frequent threat no longer exited the scene after initial onset, the “time to first fixation” measure per character was no longer relevant and was excluded. As with Experiment 1, a single fixation could be assigned to two or more ROIs if the characters visually overlaid each other at that moment. Within the critical trials, 30% of fixations were not assigned to a character, 34% were assigned to only one character and 36% to two or more characters.

Steady State Visually Evoked Potentials

SSVEPs were calculated using data collected by a BioSemi ActiveTwo Electroencephalography (EEG) system (BioSemi B.V., n.d.). The EEG system used 32 Ag-AgCI active electrodes arranged in BioSemi’s default cap configuration with six external electrodes. The 10-20 naming convention was used to identify electrode placement. Two external electrodes were placed laterally adjacent to the left and right outer canthus, one on the tip of the nose and two on the left and right mastoids. The final external electrode was placed under the left eye to record horizontal and vertical electrooculograms (EOG). All channels (electrodes) were digitally sampled at 2048 Hz, and recording occurred throughout the entire task. Electrode offset potentials were validated at the beginning of the experiment to be <50mV, referenced to the common mode sense (CMS) and driven right leg (DRL).

EEG data were processed offline using custom scripts written in Matlab (*MATLAB*) and which called EEGlab (Delorme & Makeig, 2004) functions. Raw data were down-sampled to 256 Hz, referenced against the average of all scalp channels, and the Cleanline algorithm (Mullen, 2012) was used to remove line noise. The data were additionally high-pass filtered (linear transition band 0.25 to 0.75 Hz). Data was then cleaned using the EEGlab function clean_artifacts with default settings. This function uses artifact subspace reconstruction (ASR; Kothe & Jung, 2016) to remove many of the most common artefacts from the EEG signal and to identify (and remove) bad channels. A channel was removed if it was flat for 5 seconds, had high-frequency power more than 4 standard deviations above the global average, or had a correlation less than 0.8 with its neighbours. A Butterworth low pass filter (120Hz cut-off, fourth-order) was applied to the data following ASR cleaning and prior to cleaning using independent components analysis (ICA). ICA decomposition was performed using EEGlab's "runica" command and components likely produced by eye movements were identified using the IClabel tool (Pion-Tonachini et al., 2019), subject to manual, visual inspection to confirm. All components labelled more than 50% likely to be eye movements were removed.

In order to quantify the resulting SSVEP response, we used an optimised spatial mask per individual, per trial using rhythmic entrainment source separation (RESS; Cohen & Gulbinaite, 2017). This method generates a single virtual channel, which is a weighted, linear combination of all scalp channels most sensitive to the driving frequency during a given time window using a frequency-domain Gaussian filter (full width, half maximum = 0.5Hz). This filter was also applied to the resulting virtual channel for consistency before applying the Hilbert transform to obtain a time-series of power at the driving frequency in the RESS-generated virtual channel. SSVEPs for the two cued conditions were then calculated the signal-to-noise ratio (SNR, measured in decibels) relative to the pre-trial baseline (2 seconds) in this Hilbert-transformed virtual channel. Finally, because the RESS method generates an optimized signal for the target frequency (15Hz), it is also biased to report a signal at 15Hz and will do so even when no signal is present. To control for this bias, z-scores were used to normalize the resulting signal (in dB units) in the informative and uninformative cue conditions, relative to the no-cue condition in which there was objectively no 15Hz evoking stimulus. This quantity is hereafter referred to as *normalised power*.

Mean normalized power was calculated over a 50s window extending from 25s after trial onset to 75s after onset for each trial type (training, critical, full-attention). The lower boundary (25 seconds) was based on the first time-point that all frequent threat characters were present on-screen. The upper boundary (75 seconds) was when the rare threat character most commonly exited the scene during the critical trial (this varied between participants).

Video Game Experience

Video Game Experience was captured at the end of the experiment using an adapted version of the Experiment 1 video game experience measure. As in Experiment 1 this measure was not analysed and reported in this thesis. Participants responded to the following questions:

- How would you describe your video game experience? (select one option)
    - Play regularly / Used to play a lot / Played a little in the past / Never played
- Device type(s): (select all relevant options)

    - PC, Console, Phone, Handheld device

- Game genre(s): (select all relevant options)

    - Action, Fighting, Strategy, Fantasy, Sports, Other

Other Measures

Participants were provided the same cognitive load and experiment naivety measures as Experiment 1. As in Experiment 1 only the naivety measures were analysed and reported in this thesis.

Results

Behavioural Results

Rare Threat

Behavioural responses to rare threats were the primary measure of cueing-induced inattentional blindness. Participants were classified as a noticer if they clicked on the rare threat character in the critical trial and a non-noticer if they failed to click. Across all conditions 73.4% (*n* = 69) of participants failed to click on the rare threat in the critical trial and were classified as non-noticers.

Cues Present vs Absent. The combined informative and uninformative cue conditions were compared against the no-cue condition to investigate the effect of the presence of cues, irrespective of how informative the cues were. The result is that I first consider the effects of the mere presence of a cue and then separately consider, orthogonally, the effects of the type of cue (informative, uninformative) in a subsequent analysis.

A G-test revealed a medium effect of cueing with moderate evidence that the proportion of non-noticers compared to noticers (Table 8) was greater with cues present compared to cues absent, $G(1) = 4.74$, $p = .030$, $V = 0.23$, $BF_{10} = 4.41$. An independent samples t-test on rare threat response times between cues present ($M = 11.87$, $SD = 3.39$) and cues absent ($M = 9.10$, $SD = 4.25$) during the critical trials revealed no effect; 69 missing response times were ignored. These findings show that providing cues led to increased inattentional blindness but no change in response times.

Table 8

Contingency Table Showing Noticers vs Non-Noticers for Cues Present vs Absent During Critical Trials.

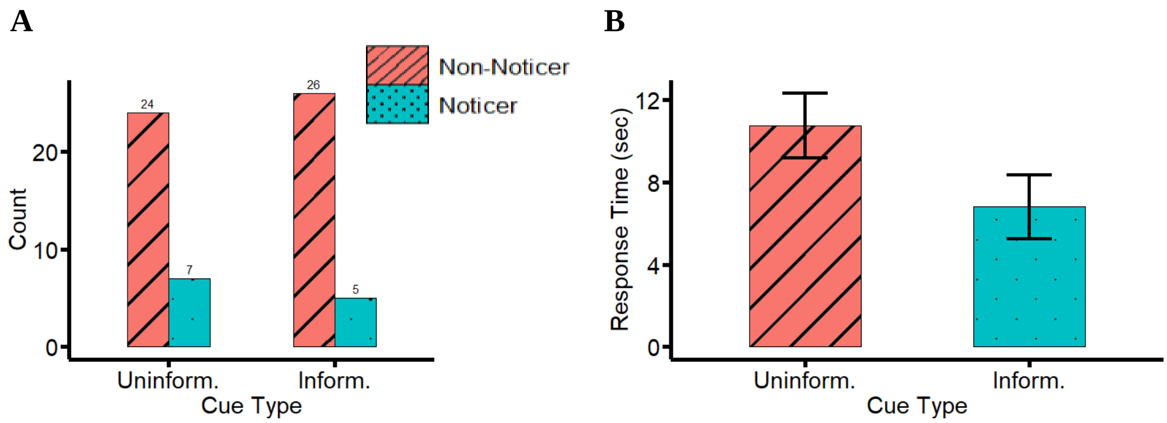
	Noticer	Non-Noticer	Total
Cues Present	12	50	62
Cues Absent	13	19	32
Total	25	69	94

Note. Noticers and Non-Noticers are classified based on clicking rare threat character during critical trials. Cues present combine both cue conditions (informative, uninformative). Absent is equal to the no-cue condition.

Type of Cue. To investigate the effect of how informative a cue was, an analysis was performed comparing the uninformative and informative cue conditions during critical trials. A G-test revealed no significant difference between the proportion of non-noticers and noticers between the three cue type conditions, $G(1) = 0.42$, $p = .520$, $V = 0.08$, $BF_{10} = 0.58$ (Figure 17A). An independent samples t-test revealed no effect of how informative a cue was on rare threat response times during the critical trials, $t(10) = 1.72$, $p = .117$, $d = 1.00$, $BF_{10} = 1.12$ (Figure 17B). Low noticer numbers (Figure 17A) make interpreting response time analysis difficult.

Figure 17

Rare Threat Behavioural Responses by Cue Type During Critical Trials.



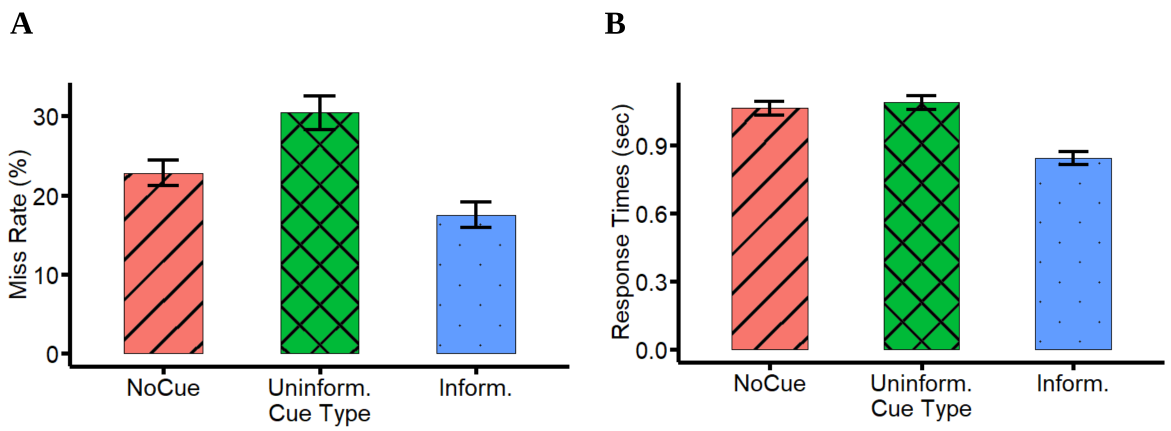
Note. (A) Noticers vs non-noticers. (B) Response times in seconds. Error bars represent standard error.

Frequent Threats

Analyses of the effect of how informative a cue was (no-cue, uninformative, informative) on the behavioural responses towards the frequent threat characters during critical trials were conducted. A oneway ANOVA showed a significant effect of cue type condition on miss rates, with extreme evidence, $F(2,91) = 12.77$, $p < .001$, $\eta^2_G = 0.22$, $BF_{10} >$ 100 (Figure 18A). Posthoc pairwise t-test results (Table 9) show that frequent threat miss rates are significantly lower in both the no-cue and informative cue conditions than in the uninformative cue condition but not significantly different when comparing the no-cue and informative conditions.

Figure 18

Frequent Threat Behavioural Responses for all Conditions During Critical Trials.



Note. (A) Miss rate (%). (B) Response times in seconds. Error bars represent standard error.

Table 9

Pairwise T-Test Results for Frequent Threat Miss Rates for all Cue Type Conditions During Critical Trials.

Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
No-cue	Uninformative	3.00(61)	.011	7.10
No-cue	Informative	2.07(61)	.124	2.34
Uninformative	Informative	5.03(60)	< .001	> 100

Note. p-adjust based on Bonferroni adjustments.

The same ANOVA was used to analyse response times. A significant effect of cue type condition on response times was found, with extreme evidence, $F(2,91) = 18.83$, $p < .001$, $\eta^2_G = 0.29$, $BF_{10} > 100$ (Figure 18B). Posthoc pairwise t-test results (Table 10) show that frequent threat response times are significantly higher in both the no-cue and uninformative cue conditions than in the informative cue condition but not significantly different when comparing the no-cue and uninformative conditions.

Table 10

Pairwise T-Test Results for Frequent Threat Response Times for all Cue Type Conditions During Critical Trials.

Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
No-cue	Uninformative	0.56(61)	1.000	0.29
No-cue	Informative	5.04(61)	< .001	> 100
Uninformative	Informative	5.56(60)	< .001	> 100

Note. p-adjust based on Bonferroni adjustments.

In summary, informative cues improved frequent threat response times compared to no cues, with a numerical but non-significant improvement in accuracy. Whilst, uninformative cues reduced frequent threat accuracy with no reduction in response times relative to no cues.

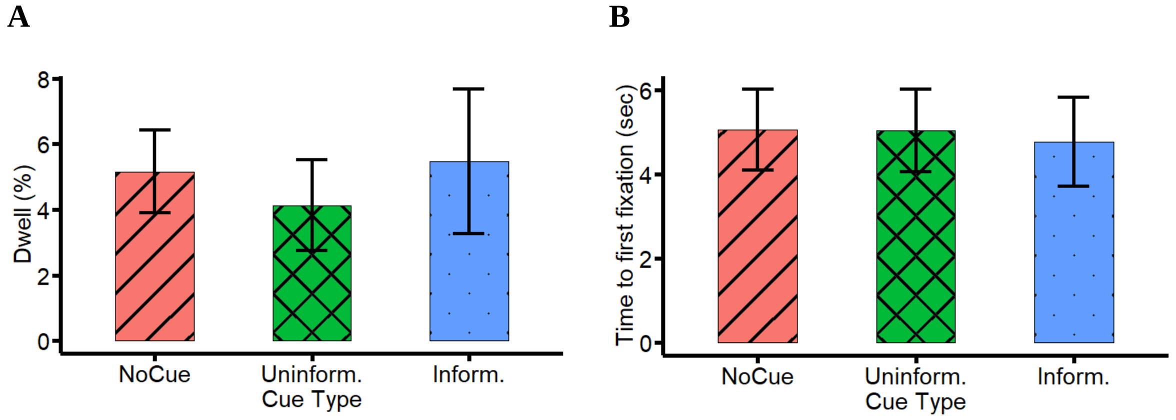
Eye Tracking

Rare Threat

The effect of cue type on time spent dwelling on rare threats during the critical trial was analysed. A oneway ANOVA found no effect of cue type condition on dwell towards the rare threat character during critical trials, $F(2,91) = 0.18$, $p = .838$, $\eta^2_G = 0.00$, $BF_{10} = 0.11$ (Figure 19A). A oneway ANOVA found no effect of cue type condition on time to first fixation towards the rare threat character during critical trials, $F(2,91) = 0.03$, $p = .974$, $\eta^2_G = 0.01$, $BF_{10} = 0.10$ (Figure 19B).

Figure 19

Eye Gaze Towards Rare Threat Characters Comparing all Conditions During Critical Trials.



Note. (A) Dwell as a proportion of time gun was visible. (B) Time to first fixation in seconds. Error bars represent standard error.

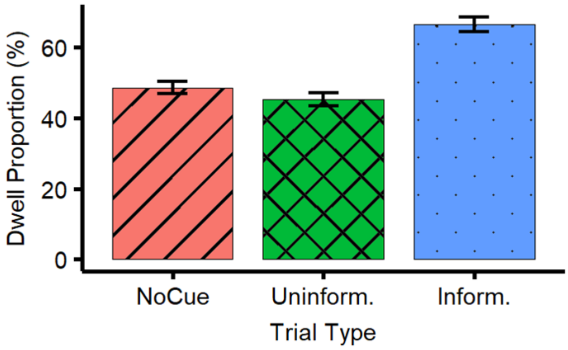
These four eye gaze outcomes combined indicate that neither a cues presence nor how informative it was influenced how fast a participant first attended to the rare threat character after onset or how much attention it garnered. This is likely due to the relatively brief period in which this data could be collected: one 17s period during only one trial.

Frequent Threat

The effect of how informative a cue was on frequent threat dwell during critical trials was analysed. A oneway ANOVA revealed an effect of cue type condition on frequent threat dwell during critical trials, with extreme evidence, $F(2,91) = 38.08$, $p < .001$, $\eta^2_G = 0.46$, $BF_{10} > 100$ (Figure 20). Posthoc pairwise t-test results (Table 11) show that the informative cue trials have a significantly higher dwell than the no-cue and uninformative cue conditions. There was no difference between the no-cue and uninformative cue trials.

Figure 20

Mean Dwell Proportion for Frequent Threat Characters During Critical Trials.



Note. Error bars represent standard error.

Table 11

Pairwise T-Test Results for Frequent Threat Dwell, all Cue Conditions During Critical Trials.

Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
No-cue	Uninformative	1.27(61)	.624	0.54
No-cue	Informative	6.89(61)	< .001	> 100
Uninformative	Informative	8.10(60)	< .001	> 100

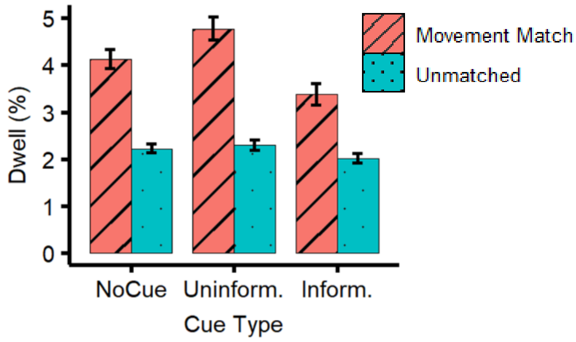
Note. p-adjust based on Bonferroni adjustments.

Distractor Type Comparison

Distractor characters were combined into two distractor type categories based on movement match with the frequent threat characters (movement matched, unmatched), and dwell was analysed for the critical trials. A mixed ANOVA was performed for dwell on distractor type (movement matched, unmatched) and cue type (no-cue, uninformative, informative; Figure 21), results showed a significant interaction, with extreme evidence, $F(2,91) = 7.36$, $p = .001$, $\eta^2_G = 0.03$, $BF_{10} > 100$. A significant within-subjects main effect of distractor type on dwell during critical trials was observed, with extreme evidence, $F(1,91) = 264.40$, $p < .001$, $\eta^2_G = 0.46$, $BF_{10} > 100$. A significant between-subjects main effect of cue type condition on dwell during critical trials was observed, with moderate evidence, $F(2,91) = 8.76$, $p < .001$, $\eta^2_G = 0.06$, $BF_{10} = 3.97$.

Figure 21

Mean Dwell for Movement Matched and Unmatched Distractor Characters During Critical Trials.



Note. Movement matched distractors replicate frequent threat character movements excluding the attack motion. Unmatched distractors never approach the friendly soldier characters. See method section for more detail. Error bars represent standard error.

A series of paired samples t-tests were used to compare the effects of cueing on individual distractor types between cue conditions (Table 12). Results for movement matched distractors revealed a significant effect with extreme evidence, with longer dwell during the uninformative cue condition than the informative cue condition. No effect was observed when comparing the no-cue to the uninformative or informative cue conditions for matched movement distractors. There was no effect observed for any unmatched distractor type comparison.

Table 12

Pairwise T-Test Results for Dwell on Distractor Characters, All Cue Conditions During Critical Trials.

Distractor Type	Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
Movement match	No-cue	Uninformative	2.05(92)	.130	1.48
Movement match	No-cue	Informative	2.35(92)	.063	3.06
Movement match	Uninformative	Informative	4.37(92)	< .001	> 100
Unmatched	No-cue	Uninformative	0.49(92)	1.000	0.28
Unmatched	No-cue	Informative	1.47(92)	.432	0.73
Unmatched	Uninformative	Informative	1.95(92)	.162	1.14

Note. p-adjust based on Bonferroni adjustments.

These results show that dwell towards distractor characters whose movement patterns *do not* resemble the frequent threat was not influenced by either the presence or not of a cue nor how informative cue where when available. However, whenever an uninformative cue was provided, dwell towards movement-matched distractors increased. The main effect of distractor type was unsurprising, as the only way for participants to distinguish between a distractor that approached a friendly soldier and a frequent threat was to watch until they commenced an attack animation or not.

Cued Characters

The ability for cues to attract attention was analysed between the two cued conditions by comparing the dwell to only cued characters, regardless of which characters were cued. An analysis that was expected to provide insight into the effect of task-relevance on attention to cues. Uninformative cues were expected to be considered less task-relevant than informative cues. The no-cue condition was excluded from this comparison as in the absence of cues there is no suitable method to select which characters to include for comparison. Frequent threat characters in the no-cue condition would offer a suitable comparison to the cued characters in the informative cue condition. However, this would generate a mismatch with the character types cued in the uninformative condition, where ~66% of cued characters are distractors and ~33% frequent threats. Leading to inappropriate comparisons between the three conditions.

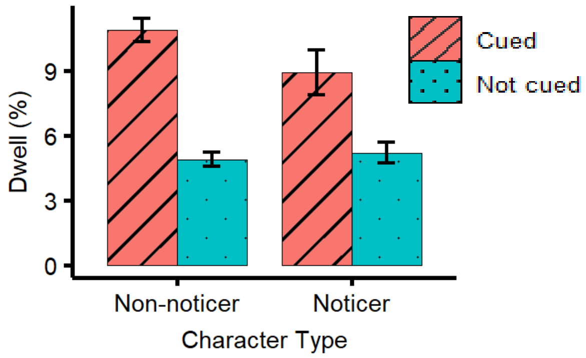
An independent samples t-test on the dwell towards cued characters between the informative (M = 13.52, SD = 2.22) and uninformative (M = 7.61, SD = 2.54) cue conditions during the critical trials revealed an effect of condition with extreme evidence, $t(59) = 9.65$, $p < .001$, $d = 2.47$, $BF_{10} > 100$.

Noticers

As the rare threat character was always uncued, any disparity in attention towards cued and uncued characters by noticer status was considered potentially informative. A mixed ANOVA was performed for dwell on character type (cued, uncued) and noticer status (noticer, non-noticer), see Figure 22. Results showed a significant within-subjects main effect of character type on dwell during critical trials, with extreme evidence, $F(1, 59) = 31.42$, $p < .001$, $\eta^2_G = 0.27$, $BF_{10} > 100$. A non-significant between-subjects main effect of noticer status on dwell during critical trials, $F(1, 59) = 1.97$, $p = .166$, $\eta^2_G = 0.01$, $BF_{10} = 0.33$. The interaction between character type and noticer status was non-significant, $F(1, 59) = 1.77$, $p = .188$, $\eta^2_G = 0.02$, $BF_{10} > 100$.

Figure 22

Dwell for Cued vs Not Cued Characters Comparing Noticers vs Non-Noticers During Critical Trials.



Note. Combines data from the uninformative and informative cue conditions. Error bars represent standard error.

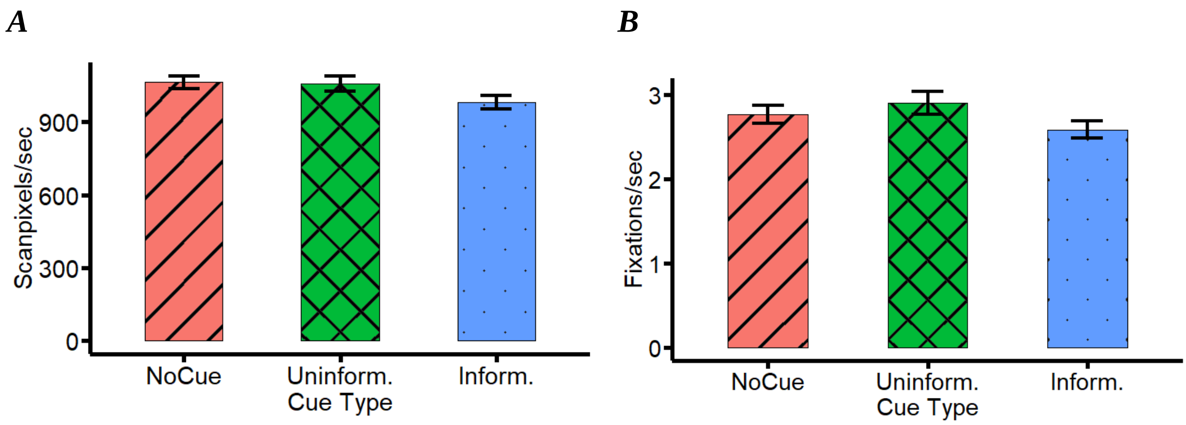
The non-significant interaction means that noticers and non-noticers did not differ in how they attended to cued and non-cued characters. A reasonable expectation was that noticers may have attended to uncued characters more than non-noticers, increasing their likelihood of noticing the uncued rare threat character. Visual inspection of Figure 22 would suggest noticers attended to cued characters less, however care needs to be taken when interpreting these results due the unbalance group sizes, 12 noticers and 50 non-noticers.

Global Search Behaviour

Global search behaviour was quantified using the scanpixels per second, fixations per second and the distance-to-nearest-friendly measures (see Experiment 1 for details concerning the rationale and calculation of these measures). A oneway ANOVA revealed no effect of cue condition on scanpixels/sec (Figure 23A), $F(2,91) = 2.51$, $p = .087$, $\eta^2_G = 0.05$, $BF_{10} = 0.70$. A oneway ANOVA revealed no effect of cue condition on fixations/sec (Figure 23B), $F(2,91) = 1.84$, $p = .165$, $\eta^2_G = 0.04$, $BF_{10} = 0.41$.

Figure 23

Global eye gaze metrics for all conditions.

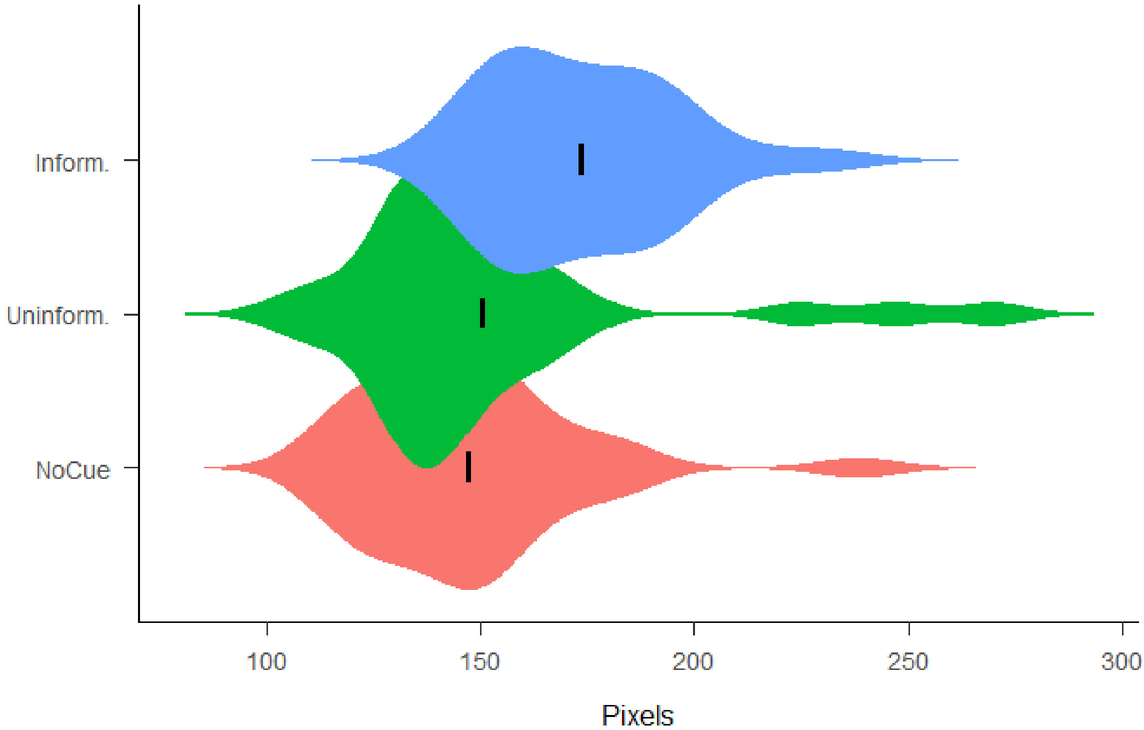


Note. (A) Scanpixels per second (B) Fixations per second. Error bars represent standard error.

Search breadth was measured via the distance to nearest friendly soldier metric. A oneway ANOVA revealed a significant effect of cue type condition on distance to nearest friendly soldier, with very strong evidence, $F(2,91) = 8.05$, $p < .001$, $\eta^2_G = 0.15$, $BF_{10} = 49.75$ (Figure 24). Posthoc pairwise t-tests (Table 13) reveal that those in the informative cue condition had a greater search breadth than those in the uninformative cue or no cue conditions, and that there was no significant search breadth difference between the uninformative and no cue conditions. Participants spread their gaze similarly when either no cue or an uninformative cue was provided. However, providing an informative cue resulted in consistently broader search patterns.

Figure 24

Violin Plot of Distance in Pixels of Fixations From Nearest Friendly Soldier Character by Cue Type During Critical Trials.



Note. Bars show mean distance in pixels.

Table 13

Pairwise T-Test Results for Distance to Nearest Friendly, all Cue Conditions During Critical Trials.

Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
No-cue	Uninformative	0.47(61)	1.000	0.28
No-cue	Informative	3.70(61)	.001	> 100
Uninformative	Informative	3.20(60)	.006	11.15

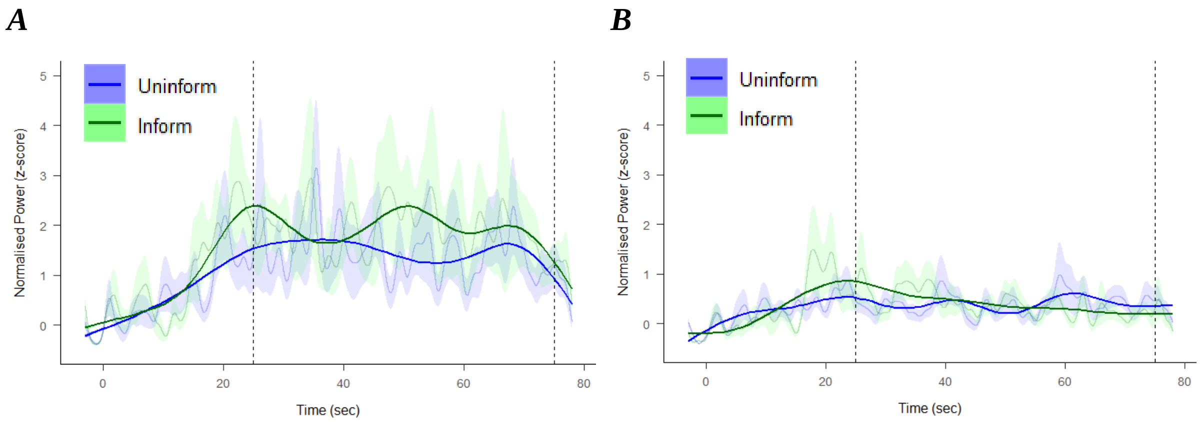
Note. p-adjust based on Bonferroni adjustments.

SSVEP

The time series of the power-normalised SSVEP signal for the informative and uninformative cue conditions across the course of the critical and full-attention trials is shown in (Figure 25). Vertical dashed lines indicate the measurement window. Consistent with the initial hypothesis, power appeared greater for informative cues than uninformative cues in the critical trial, but then global power dropped, and no distinction was apparent in the full-attention condition. However, inferential analyses did not support this interpretation. An independent samples t-test on the mean value of these z-scores between the informative and uninformative cue conditions revealed no effect of cueing in either the critical trial, $t(59) = 0.48$, $p = .632$, $d = 0.12$, $BF_{10} = 0.29$ (Figure 25A), or the full-attention trial, $t(57) = 0.02$, $p = .985$, $d = 0.01$, $BF_{10} = 0.27$ (Figure 25B). Overall, no evidence for an effect of cueing type on SSVEP power was observed.

Figure 25

SSVEP Normalised Power Comparison Between Informative and Uninformative Cue Conditions



Note. (A) Critical trials. (B) Full-attention trials. Bold lines represent smoothed results with raw values underneath. Vertical dotted lines indicate the range of values used for inferential tests. See the measures section above for further details. Error bars represent standard error.

Full-Attention Trials

The full-attention trials provided an opportunity to investigate if the rare threat character was plainly visible when participants were unburdened by the demanding frequent threat search task. Expectancy of the rare threat was also expected to be higher in the full-attention trials due to a combination of task instructions and potentially having noticed the rare threat during the critical trial. Full-attention trial rare threat character data was grouped into a 2 (Hits, Misses) x 3 (no-cue, uninformative, informative) table (Table 14). As expected, detection rates were high (92.55%), averaged across all groups. A G-test showed that the proportions of hits and misses between cue-type conditions during full-attention trials were not significantly different, $G(2) = 2.00$, $p = .300$, $V = .15$, $BF_{10} = 0.22$.

Table 14

Contingency Table Showing Rare Threat Character Hits and Misses for all Cue Type Conditions During Full-attention Trials.

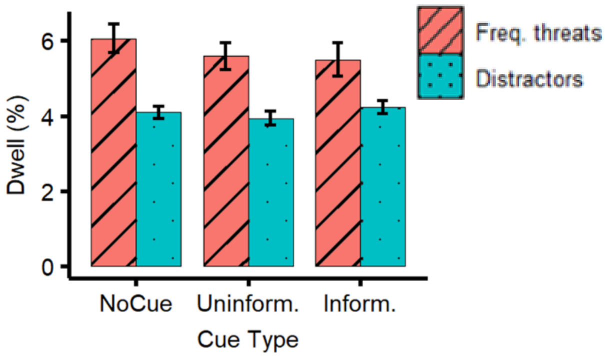
	No-cue	Uninformative	Informative	Total
Hits	30	30	27	87
Misses	2	1	4	7
Total	32	31	31	94

A mixed ANOVA was performed to evaluate the effect of cue type condition (no-cue, uninformative, informative) and character type (distractor, frequent threat) on dwell (Figure 26) during the full-attention trials. The within-subjects main effect for character type was significant, with extreme evidence that participants dwelled longer on frequent threats than distractors, $F(1, 90) = 45.90$, $p < .001$, $\eta^2_G = 0.19$, $BF_{10} > 100$. The between-subjects main effect of cue type on dwell was not significant, $F(2, 90) = 0.54$, $p = .587$, $\eta^2_G = 0.01$, $BF_{10} = 0.09$. The interaction between character and cue types on dwell was not significant, $F(2, 90) = 0.72$, $p = .488$, $\eta^2_G = 0.01$, $BF_{10} = 0.42$.

Turning to the related question of whether people gazed more at characters overlaid by cues. An independent samples t-test showed no significant effect on dwell towards cued characters in the uninformative ($M = 5.44$, $SD = 2.19$) vs informative ($M = 5.50$, $SD = 2.42$) cue type conditions, $t(59) = 0.10$, $p = .923$, $d = 0.03$, $BF_{10} = 0.26$.

Figure 26

Dwell by Cue Type for Frequent Threat and All Distractor Characters During Full-Attention Trials.



Note. Distractors include all distractor types. Error bars represent standard error.

These results indicate that during the full-attention trials when participants are unburdened by the primary search task the rare threat character was plainly visible and that any effects of cue type on gaze towards frequent threats, cued characters or distractor characters observed in the critical trials largely disappeared. However, frequent threat characters still garnered more attention than distractors despite no longer being task-relevant.

Inattentional Blindness Awareness

Awareness of inattentional blindness and the gorilla experiment (Simons & Chabris, 1999) were queried post-experiment. In brief, 20% of participants reported awareness of inattentional blindness, and 36% were aware of the gorilla experiment. No participants indicated associating this experiment with inattentional blindness or the gorilla experiment.

Discussion

Informative cueing in Experiment 1 appeared to improve the detection of cued objects with little to no cost to detecting the uncued, unexpected, rare threat. The present data reveal that this cost/benefit relationship is more nuanced. The present experiment introduced an uninformative cue condition, limiting the rare threat to one appearance in each critical and full-attention trial. The latter change necessitated a within-subjects rather than between-subjects inattentional blindness comparison between cue-type conditions, reducing statistical power, which compelled a larger sample size. However, the primary benefit of the present design was improved purity of measurement. When the rare threat appeared in the critical trial, it was unambiguously unexpected, just as in the protocol pioneered by Mack, Rock, and colleagues (see Mack & Rock, 1998 for an overview). Under these conditions, an effect of cue-type on inattentional blindness was observed.

Did Cueing Work?

The cueing implementations were effective. The informative cue condition improved detection rates and speed for the frequent threat characters when validly cued relative to the uninformative and no-cue conditions. Further, there was suggestive evidence that the introduction of an uninformative cue, which included invalidly cued distractor characters, actively impaired performance relative to the no cue control condition. Specifically, the uninformative cue condition produced higher miss rates but similar response times to the no-cue condition for frequent threat characters.

Rare Threat Behavioural Responses

Overall, ~73% of participants were classified as non-noticers, having experienced inattentional blindness. Noticing rates were significantly reduced under conditions where a cue (informative or uninformative) was on the screen relative to the cue absent condition.

Interestingly, there was a marginally significant finding in response time in the opposite direction. If participants noticed the rare threat in the cued conditions, which they were less likely to do, they tended to notice it faster. These paradoxical findings are explored in the General Discussion chapter. The utility of the cue (informative or uninformative) did not appear to impact detection accuracy or speed. Finally, it is important to note that many of these comparisons between noticers and non-noticers were necessarily based on small and uneven group sizes; the experimenter has no direct control over noticing rates. Noticing rates were only ~16% (n = 5) of participants in the informative condition and ~22% (n = 7) in the uninformative, compared to ~40% (n = 13) for the no-cue condition.

Attentional Tunnelling

Attentional tunnelling towards cued characters (informative and uninformative) was predicted to increase inattentional blindness towards uncued rare threats. This tunnelling should have manifested as attentional resources drawn towards cued characters and away from uncued ones, including the rare threat character. It was expected that this bias towards cued characters would have been stronger when cues were informative, resulting in uncued characters (including the rare threat) being insufficiently attended to and processed to determine whether these characters were a threat. This hypothesis has several separable components. The evidence for each component is considered in turn, starting with evidence bearing upon the presence of attentional tunnelling.

Evidence for attentional tunnelling was only convincingly evident in the informative cue condition. Cued characters were attended to longer than uncued characters in the informative than the uninformative cue condition. This bias towards validly cued characters translated into improved frequent threat detection speed and accuracy for the informative cued condition.

Within the informative cue condition, there is strong evidence that the highly task-relevant cues attracted and held attention. Compared to the uninformative cue condition, the informative cue condition shows nearly double the dwell towards cued characters, increased dwell towards frequent threat characters and less attention towards distractors. Wider search breadth within the informative condition also suggests participants tracked cued characters more broadly across the scene than in the other conditions. Improved behavioural responses towards frequent threat characters suggest this increased attention improved performance on the primary search task, potentially at the cost of the search for the uncued rare threat characters. Outcomes often associated with attentional tunnelling and overuse of a decision aid.

In contrast, the uninformative cue condition shows similar dwell towards rare threat and distractor characters and equivalent frequent threat response times compared to the no-cue condition, and worse frequent threat accuracy when compared to the informative cue condition. In summary, results more similar to the no-cue than the informative cue condition and outcomes that would be expected if the task became more difficult.

The second consideration was that the cue's task relevance, not its contrast, brightness or salience would produce attentional tunnelling. On this assumption, attentional tunnelling would be displayed if two gaze bias patterns were observed. Firstly, that gaze would be biased towards informative cues over uninformative cues in the critical trial. Finally, that this gaze bias to informative cues during the critical trial markedly reduces in the full-attention trial when those same visual cues are no longer task-relevant. Indeed, this was precisely the patterns observed. The perceptual features of the cues remained the same on both the critical and full-attention trials. Then, with the cue's relevance to the search task removed during the full-attention trials, the bias towards the cues diminished, revealing that cues were prioritised by their relevance, not their physical attributes.

Visual Clutter

Perhaps the increased inattentional blindness rate is better understood as an effect of visual clutter. Cues induced attentional tunnelling in the informative cue condition, and there was relatively little evidence of this in the uninformative cue condition. Yet detection of the rare threat was reduced by the mere presence of cues, irrespective of whether those cues were informative or uninformative.

Increasing visual clutter can increase search difficulty. For example, adding visual cues to the scene increases visual clutter, making detecting the uncued, rare threat more difficult. Henderson et al. (2009) showed that when searching in real-world scenes, an increase in visual clutter is akin to an increase in set size. Increases in set size reliably increase serial, but not parallel, search difficulty (Horowitz & Wolfe, 1998; Wolfe, 2007).

In some important respects, searching for a rare threat event can be considered an instance of serial search. The rare threat is uncued and is far less frequent than the visually similar characters it is presented amongst. Adding the rifle to the character could induce a popout effect (and thus induce parallel search). Yet this seems unlikely in the context of the critical trial as it was detected by fewer than 30% of participants in the 17s in which it was visible. So if detecting the rare threat relies on serial search and serial search performance reduces as set size increases (Henderson et al., 2009), then anything that increases set size will reduce rare threat search performance. With visual clutter acting as a proxy for set size (Henderson et al., 2009), adding visual clutter to the scene via visual cues (informative or uninformative) will likely reduce rare threat search performance, outcomes that were observed in the critical trials.

In an apparent contradiction, this effect of set size (via visual cues) did not occur in the full-attention trials. Miss rates for the rare threat in the final, full-attention trial were uniformly low across all three conditions. One possibility is that the impact of visual clutter on performance was reduced when participants were no longer required to perform the primary search task. However, this explanation is speculative and potentially at odds with the definition of visual clutter: clutter increases perceptual complexity. It thus should impact any task involving perceptual processing of the scene. Instead, an alternative explanation in terms of expectancy was explored.

Expectancy

The final, full-attention trial is a commonly used control trial in inattentional blindness experiments (Mack & Rock, 1998). It is generally understood to act as a control for the effects of task load. That is, it is presumed that the near 100% noticing rate in the full-attention trial was due to decreased task difficulty due to the frequent threat search task being eliminated. Indeed, this logic underlies its name: full attention.

Yet, the full-attention trial differs from the critical trial in another important respect. Expectancy of the rare threat character also increases at the commencement of the full-attention trial. Participants are instructed to no longer search for frequent threat characters and only search for other threats alongside a reminder of the task instructions - “search for any weapon”. Furthermore, approximately 27% of participants clicked on the rare threat in the critical trials. With the reminder combined with the possibility of previously noticing the rare threat, it is reasonable to assume that this would increase participants’ expectancy of the rare threat character. Expectancy has a strong, positive effect on detecting the rare threat in inattentional blindness studies (Mack & Rock, 1998).

This increased expectancy might explain why cueing selectively impacted detection rates of the rare threat during the critical trial (akin to an effect of visual clutter) but not on the subsequent, full-attention trial. On the full-attention trial, participants in all conditions were essentially instructed to ignore the characters they had previously been monitoring and actively search for a new event. Such top-down, endogenous search for the rare threat, driven largely by a step-change in expectancy, may override any effect of visual clutter evident in the critical trial.

Inattentional Blindness Summary

The preceding analysis did not derive a singular definitive account for the low noticing rates in the two types of cued trials. Low noticing rates in the informative cue condition might have been the product of attentional tunnelling or visual clutter or their combination. By contrast, low noticing rates in the uninformative condition were less likely to be the product of attentional tunnelling, as the evidence of tunnelling in this group was meagre. Instead, low noticing in this group was more likely due to the uninformative cues constituting visual clutter sufficient to disrupt the search. So, while the uninformative and informative condition noticing rates were similar, the mechanisms underlying each may differ. This accords with a recent experimental series investigating inattentional blindness, in which small changes in the stimulus set appeared to invoke distinct attentional strategies in participants, resulting in qualitatively different search patterns across superficially similar inattentional blindness tasks (Hutchinson et al., 2023).

Effect of Gaze on Noticing

When eye tracking is available within inattentional blindness research, it is common to investigate if noticing was conditional on fixating the rare/unexpected stimuli (Beanland & Pammer, 2010; Hutchinson et al., 2023; Richards et al., 2012). The present study collected much of the relevant data to perform these analyses. However, the present experiment design required noticers to target and click on the rare threat character while it was onscreen. This behavioural response task required fixating on the rare threat character, thereby making it difficult to determine gaze behaviour attributable to search as distinct from gaze behaviour attributable to generating a response amongst the noticers. I therefore decided not to perform these analyses as it was unlikely to contribute reliable knowledge to the existing literature.

Inattentional Blindness Awareness

Awareness of inattentional blindness and the gorilla experiment (Simons & Chabris, 1999) were queried post-experiment, with 20% and 36%, respectively, reporting awareness. I concluded the participant pool was sufficiently naïve to both as no one reported associating the current experiment with either until after these questions.

SSVEP

Using the SSVEP technique in this project was an exploratory exercise. No prior research appears to attempt SSVEP in such a dynamic and visually complex scenario as the present research. It was considered an attractive technique as SSVEP is reportedly more robust to head and eye movements than other EEG measures (Norcia et al., 2015). Suggesting an opportunity to determine if such a technique could be sensitive enough within our experiment to be of value.

The challenges to SSVEP within the present experiment were numerous. Of primary concern was that nine frequency-tagged (tagged) stimuli traversed the scene in a pseudo-random manner at all times, and participants were required to overtly attend to either some or all of these tagged stimuli to complete the primary task. The requirement to overtly attend to frequency-tagged characters was likely to generate SSVEP signals of a magnitude that they would overpower any other SSVEP signals, such as covert attention. Previous researchers have reported that overt signals can be up to ten times the power of covert attention signals (Walter et al., 2012). An outcome that would make the SSVEP results analogous to the existing eye gaze results and, therefore, superfluous. Further complications included that participants were unrestrained, needed to actively search the scene presented and used the computer mouse to click on stimuli. Introducing the potential for more muscle artifacts than is desirable for EEG data recording (Luck & Kappenman, 2017).

Analysis of the SSVEP results demonstrated that despite the above limitations, SSVEP signals were apparent at the tagged frequency (15 Hz). However, this difference was not statistically significant despite visual appearances of a difference in attention between conditions (Figure 25) during the critical trial. Thus, the present analyses could not distinguish between the informative and the uninformative cueing conditions when the rare threat was present. However, the measure did appear to be sensitive to task relevance, with the full-attention control trial reduced to <30% of that observed in the critical trials. This reduction reflects participants attending less to the cues during these full-attention trials when the cues were task-irrelevant.

The conclusion from the critical versus full-attention trial comparison is that it allays one concern of using SSVEP within this context: that the design of the scene and cues would result in ceiling levels of SSVEP regardless of what participants were attending to. If this were to occur, no reduction in SSVEP power from the critical trials to the full-attention trials would have been observed.

In summary, despite some promising signs, there were no indications that SSVEP was sensitive to shifts in covert attention, a primary motivator for its inclusion. Thus, it failed to provide any knowledge beyond that already obtained from the eye gaze results.

Conclusion

The present experiment addressed stimulus and design limitations that may have hampered experimental sensitivity to a cueing effect in Experiment 1. The experiment additionally included a neural measure of attention: SSVEP.

An increase in inattentional blindness was observed when cues were provided. Two possible explanations for the observed cueing-induced blindness were considered, one based on attentional tunnelling and one on visual clutter. Only the clutter explanation held for both cueing conditions, as there was limited evidence of tunnelling in the uninformative cue condition. Thus, on balance, informative cueing improved performance (reaction time and accuracy) for cued threats. However, the mere presence of these cues reduced sensitivity to unexpected, rare threats. The cost/benefit calculation for uninformative cues was clearer: uninformative cues reduced detection performance for frequent threats and reduced capacity to detect the rare threat.

To optimize the manipulation of cue utility, the present experiment compared artificially accurate (perfectly valid and reliable) cues and perfectly uninformative (zero contingency) cues. Yet even under these optimistic assumptions, neither scenario was without cost. Discussion of generalising the cost/benefits identified above to more likely rates of validity and reliability has been deferred to the General Discussion. In the next chapter, the generalisability of the present findings is explored by deploying the same task in virtual reality. Bringing the experimental cueing paradigm closer to the phenomenon we sought to model: augmented reality in head-mounted displays on dismounted soldiers.

Chapter 4: Experiment 3

Introduction

Some degree of inattentional blindness was observed across all conditions in the previous two experiments. The within-subjects design in Experiment 1 did not observe any effect of cueing, but this was likely due to the strong role expectancy played in a task in which the rare threat occurred six times across the critical trials (and a further six in the full-attention trials). This hypothesis was supported by behavioural data; people were much more likely to observe rare threats that occurred late in the critical trials, than earlier in the trials (85% on sixth appearance compared to 37% on first). After people notice a rare threat once, they were more likely to achieve detections subsequently.

In Experiment 2 an effect of cueing on inattentional blindness was observed broadly across both informative and uninformative cues. This pattern was not supported by either of the initial hypotheses: valid cueing was expected to decrease inattentional blindness by reducing cognitive load, or to increase inattentional blindness via attentional tunnelling towards cues. Neither of these mechanisms ought to have induced a comparable effect in the uninformative cues, as they did not support attentional tunnelling (defined as preferential gaze and speeded frequent threat detection) nor did they reduce load (if anything, they impaired performance). Yet cueing generally, informative and uninformative alike, produced more inattentional blindness.

Consequently, we hypothesized that serial search for the unexpected rare threat would be interrupted by a visual clutter, or crowding, effect (Levi, 2008; van den Berg et al., 2009) induced by the unordered, dynamic, salient visual elements added to the display (the cues). However, even this explanation is incomplete. For example, inattentional blindness may have been achieved in different ways for the two cueing conditions (see also Hutchinson et al., 2023), such that attentional tunnelling was additionally present in the informative cueing condition, but not the uninformative condition, and may have contributed to the inattentional blindness rates observed. Further, the impacts of visual clutter were only evident when people were tasked with performing concurrent search for the frequent threats. When this task was removed in the full-attention condition, and people's expectancy of the rare threat was increased, cueing exerted no influence on rare threat detection. In brief, none of the accounts considered offers a sufficient and unambiguous account of all data observed. The impact of cues on inattentional blindness in a given task may depend on the relative contribution of several interrelated factors: expectancy, attentional tunnelling, task difficulty and visual clutter.

One might, therefore, expect the impact of cueing on inattentional blindness to be specific to the format in which the search task is presented. The present experiment explored the same hypotheses in the same scene, using a virtual reality head-mounted display to more closely model the experience of augmented reality cues in a dismounted soldier in the field. There are reasons to believe the presentation format used in Experiments 1 and 2 (a video game on a computer monitor) may have increased the contribution of some attributes impacting inattentional blindness more than others.

Specifically, presenting the experiment on a 2D computer monitor facilitates visual clutter. The experiment simulates a military overwatch where a soldier looks over a market square. Despite the scene being rendered in a modern, high quality gaming engine, the nature of illustrating depth in a game scene presented on a computer monitor meant that participants experienced a somewhat down-scaled and flattened version of the 3D scene. For example, stimuli that were several meters apart in the 3D world may only be a few pixels apart, or even overlap, in the isometric projection shown on the computer screen.

This situation mostly arises out of the complex geometry involved in calibrating the in-scene virtual camera and the actual display size and distance in the user's environment. In our particular case, a default in-scene camera (90 degrees) was projected onto a 24-inch monitor viewed at 60cm. This compresses 90 horizontal degrees of in-scene vision to approximately 45 degrees in the visual field of the seated user. The user can see more scene in the game world than they would be able to in their real world, leading to some compression in geometry. To the extent this compression occurred it would elicit more visual clutter (or perceptual crowding) than the same scene might evoke if it was viewed by a person within the actual scene. Such effects are reduced in virtual reality. The HTC Vive headset used in this experiment offers 110 degrees field of view, which more closely aligns with the Unity systems 90 degree in-scene virtual camera. Avoiding the scene compression experienced when the same scene is viewed on a desktop monitor. The scene in VR is therefore projected over more of the user's visual field, resulting in a user experience closer to one-to-one with what an actual operator in the scene might experience. For a similar reason, it results in less variability in the projection parameters between users too (retina-to-display distance is more tightly controlled in a VR headset than in free screen viewing, as the headset itself physically constrains it). Because there was less compression of the visual field in the VR display format, it was anticipated that the effects of visual clutter would be reduced relative to Experiment 2, thereby allowing more sensitivity to any effects of attentional tunnelling.

Putting aside the potential impact of cueing, the more general question of whether display format impacts inattentional blindness has been explored previously. For example, Chabris et al. (2011) investigated inattentional blindness in a closer to real-world setting modelled on the experiences of an American police officer accused of failing to report a crime. This officer was found guilty of various specific charges, and this ruling hinged on the observation that the officer either failed to notice the assault of a colleague despite running directly past the assault whilst pursuing a suspect or he noticed the assault and intentionally failed to render assistance or report the assault (Lehr, 2009, as cited in Chabris et al., 2011). Jurors did not believe his claim that he did not notice the assault as he ran past. Chabris et al. (2011) simulated the officer's foot chase by having participants chase a confederate around a university campus, including running past a nearby simulated assault. Results showed that only 56% of participants noticed the assault during the daytime (fewer under degraded or challenging viewing conditions), despite the assault occurring within sight of their running path. The study shows that inattentional blindness can be replicated in real-world experiments with similarly low noticing rates to lab or screen based inattentional blindness research (see meta-analysis in Ekelund et al., 2022). Inattentional blindness is not a mere artefact of screen-based projections or the use of artificial stimuli. Further, there is reason to believe that inattentional blindness rates can be at least as high in real world situations in which participants are placed in the scene, than in sterile, safe lab-based testing environments.

The Chabris et al. (2011) study is a good example of ecologically valid research. However, there is an inherent trade-off between optimizing for ecological validity and optimizing for experimental control (internal validity). Virtual reality offers an important middle ground, because it offers the participant a more immersive experience than when performing the same task on a computer monitor (Brookes et al., 2020; Kisker et al., 2019) and yet the experiment can retain tight control over the scene. For instance, stimulus positions and timing can be controlled, biometric data (EEG and gaze) can be readily recorded and time-locked to stimulus presentations, and the scene can be easily repeated thereby allowing the classic block and trial experimental design elements.

Nevertheless, virtual reality research taking place in video game worlds populated by autonomous agents cannot offer the same precision of control as lab-based tasks using a single, precisely controlled stimulus set. For example, in the present Unity scene, character pathfinding meant that while we could precisely measure stimulus onsets and offsets, we could not directly control them. Classic inattentional blindness tasks using geometric stimuli do not suffer the same limitation. Relatedly, in classic screen-based tasks, the whole scene remains constantly within the participants' field of view. Within VR tasks, the programming only decides what is "potentially" visible. Participants are free to move their heads, and as they do, what they can see changes. I used the phrase potentially visible to represent that the experimenter/programmer is no longer completely in control of what the participants see.

Schöne et al. (2020) compared the presence of inattentional blindness in screen-based and VR experiments. Schöne et al. (2020) replicated the iconic Simons and Chabris (1999) inattentional blindness experiment, where a person in a gorilla suit walks through a basketball game. They created a three-dimensional (3D) video that closely replicated the original study's video. They then had 19 participants view this video via a typical computer screen and 22 participants via a virtual reality (VR) headset. The expectation was that the VR headset would increase the realism of the experiment compared to the computer screen, therefore being a closer approximation to a real-world scenario. Their results show a significant improvement in noticing rates for those in the VR condition (~70%) compared to those in the screen-based condition (~33%). Assuming that VR is a closer approximation to real life than viewing a video on a monitor, then the magnitude of change between the screen and VR experiments suggests that inferences from screen-based inattentional blindness studies may overstate the extent of inattentional blindness in real-world tasks.

Schöne et al. (2020) concluded that the immersive VR experience increased the gorilla's salience, leading to increased self-relevance and potentially a fear or similar response. They speculated that the gorilla was perceived as spatially close in the VR scene (relative to on a computer monitor), and of a similar physical size to the viewer. In turn, these perceptual features increased the perception of imminent risk or threat (of having an undetected gorilla nearby), which then led to higher detection rates. While this reasoning appears slightly circular – the character is perceived as a threat and thereby more likely to be perceived – it remains possible that low level perceptual characteristics associated with physical proximity and known to moderate search efficiency (two-dimensional expansion or “looming;” Takeuchi, 1997) may have disproportionately captured attention in the VR format relative to the screen format, resulting in higher VR detection rates.

In contrast to Schöne et al. (2020), the present experiment is constructed from the perspective of a viewer perched on a building above a market scene, with the characters in the scene (the search array) presented below the player’s avatar (approx. 30° depression). This means the characters never enter close proximity to the participant. Additionally, any perceived threat is towards others and not the participant, and the threat remains approximately the same distance from the viewer throughout all trials. As a result, any perceived stimuli size or self-relevance changes ascribable to VR would not be on the same scale as evident in Schöne et al. (2020). But what may be of relevance is the potential for more realistic depth perception due to the immersive experience. The Schöne et al. (2020) scene was shot with a multi-lens commercial camera designed to generate an impression of depth in VR presentation formats by overlaying adjacent images. However, our scene was fully modelled in three dimensions and only at runtime was it projected into a format suitable for the VR display, thereby allowing for more comprehensive depth information and likely a better sense of immersion. For example, perception of immersion increases when participants can choose their viewpoint (Lo & Lai, 2023). In our animated scene, small translations of head position forward, backwards and side-to-side are readily acknowledged and viewpoint is adjusted. In scenes generated by static multi-lens cameras, often only rotation is acknowledged, thereby limiting immersion.

Of specific interest is the influence that improved spatial realism may have on the search strategies employed by participants. For example, the limited display size of screen-based studies may allow participants to select a more uncomplicated strategy that minimises head and eye movements, and instead leverages their peripheral vision. In contrast, the more immersive and spatially realistic experience within VR may encourage a more active strategy with larger head and eye movements required to achieve suitable coverage of the search zone. Any potential changes in search strategy become important as generalisability suffers if the outcomes observed were due to a search strategy that would not be used in real-world contexts.

In addition to the commonly discussed immersion of VR, there are other factors to consider when comparing screen-based and VR studies. Of particular note for the present experiment is the targeting device. In the prior two studies, screen-based participants used a computer mouse to move the cursor over the stimuli and then click. Whereas within VR, they will use a hand controller (see Figure 27Method section), instead of moving the cursor, they have a laser beam to sight the stimuli and a trigger style button to click. Anecdotal feedback during pilot testing indicated that this hand controller was significantly faster and easier to target than a computer mouse. The subsequent decreased attentional demands of targeting within VR may further influence how they search for and attend to objects with the experiment.

This experiment aimed to replicate the previous experiment's design as closely as practical within the VR environment. The underlying task, trial, and block structures remained the same. The game scene and characters required minimal changes when adapting to the VR environment. Primary differences related to how eye-tracking and EEG were captured and analysed. Refer to the method section below for specifics.

The predictions for this experiment were based on the findings from Experiment 2. Of note is that critiquing or contrasting screen versus VR experiments is of secondary interest. The primary focus remains on examining whether, and how, AR cues help search, and where they may hinder it. Moving the experiment into VR provides an opportunity to investigate the impact of cueing on search in a more spatially realistic environment. Some inter-display comparisons will be inevitable and predominantly included in the general discussion chapter of this thesis.

The primary hypothesis predicts that based on the Experiment 2 findings, inattentional blindness patterns will be similar between the informative and uninformative cue conditions. It is worth noting however that there are now two competing viable hypotheses. The hypothesis selected presumes that attentional tunnelling towards the cues (either informative or uninformative) during the critical trials was the determinate of inattentional blindness in Experiment 2. Whilst the alternate hypothesis presumes that the determinants of inattentional blindness differed between the two conditions; with attentional tunnelling accounting for the informative cue results and visual clutter accounting for the uninformative cue results. In this later hypothesis, the increased spatial realism of VR would be expected to reduce inattentional blindness rates for the uninformative cue condition only.

Hypotheses

- Rare threat detection performance was predicted to be lower in the informative and uninformative cue conditions compared to the no-cue condition, with higher miss rates and response times.
- Frequent threat search performance was predicted to be higher in the informative cue condition compared to the uninformative and no-cue conditions, with lower miss rates and faster response times.
- Overt attention (dwell) towards rare threat characters was predicted to be higher in the no-cue condition than in the uninformative cue condition, which was further predicted to be higher than in the informative cue condition.

- Overt attention (dwell) towards frequent threat characters was predicted to be higher in the informative cue condition than in the uninformative cue condition, which was further predicted to be higher than in the no-cue condition.
- SSVEP power was predicted to be higher in the informative cue than in the uninformative cue condition.

Method

Participants

Thirty-nine participants from the Flinders University First Year Psychology undergraduate participant pool (29 Female, 10 Male; mean age 24.1 years, *SD* = 9.31) participated in exchange for course credit. All participants gave informed consent and during pre-screening reported normal or corrected-to-normal vision and normal colour vision. This research complied with the tenets of the Declaration of Helsinki and was approved by the Social and Behavioural Research Ethics Committee at Flinders University (Project ID: 2569).

COVID-19 related restrictions in place throughout the experimental period largely dictated access to participants and therefore final participant numbers. As this study largely replicated Experiment 2 the power analysis from Experiment 2 was considered adequate, with a suggested sample size of 88 to attain 80% power to detect a medium effect.

Stimulus and Materials

The experiment was an adaption of the experiment used in Experiment 2 in the previous chapter. The only changes required were those needed to run the experiment via a VR headset. Participants used an HTC Vive Virtual Reality headset (*VIVE VR System*, 2022) operating at a 90 Hz refresh rate. A Tobii Pro VR integration kit (Tobii AB, 2015) providing 3-dimensional (3D) eye tracking at a 120 Hz sample rate was retrofitted to the HTC Vive headset. Experimenters could monitor the participant's experiences via a monitor providing a 2D display of the VR scene. A BioSemi ActiveTwo Electroencephalography (EEG) system collected data to enable SSVEP analysis, see measures section for more detail. See Figure 27 for an example of the EEG and VR equipment setup.

Figure 27

EEG Cap and VR Headset and Hand Controller as Fitted to Participant During a Trial

The experiment was equivalent to the video game representation of a military overwatch scenario used for Experiment 2 (Figure 28). Participants looked down on a market square where various characters moved around the scene performing multiple interactions and animations. Unity used a world-space referencing system to map the 3D model of the video game scene into VR. Objects such as buildings and characters were positioned within the game scene using (x, y, z) references relative to the Unity world origin (x=0, y=0, z=0). The Unity system logged the world-space coordinates of every character within the scene at a 90Hz sample rate. See the Unity online documentation for a complete overview of the Unity system (*Unity*, 2020). The Tobii eye-tracking system logged gaze vectors for both the left and right eyes, representing gaze vectors as unit vectors using the Unity world-coordinate system for the origin and direction of the gaze.

Character design and behaviour remained the same as in Experiment 2; refer to Figure 28 for an example trial snapshot and the previous chapter for detailed character design features. Similar to Experiment 2, this experiment is a between-subjects design with the experiment split into three trial types (training, critical, full-attention) with participants alternately allocated to one of three conditions no-cue (n = 13), informative cue (n = 13) or, uninformative cue (n = 13). The design of these conditions matched those from Experiment 2.

Figure 28

Example of the experiment during a cued trial.

Note. Image contrast adjusted for publication purposes.

Procedure

Commencement and instructions to participants remained similar to Experiment 2, except for the following key elements. The experimenter set up and adjusted the VR headset and EEG cap/electrodes for the participant. Participants were seated in a chair, and due to the hard-wired VR/EEG connections, instructed to minimise movements such as pivoting. They were not otherwise constrained and could still rotate the chair or move their head to alter the view within the VR headset. Experimenters instructed participants on the use of the VR hand-controller. Participants received the same condition specific instructions as Experiment 2, repeated here in Table 15 for convenience.

Table 15

Specific Instructions by Condition.

Cue Condition	Instructions
No-cue	You are in the no-cue condition. No target detection aids will be present and you will have to identify threats on your own.
Uninformative	You are in the uninformative cue condition. Red cues will appear over some characters who will approach a soldier, pull out a knife and attack. But cues may also appear over characters who are not a threat. Therefore, you must try and ignore the cues.
Informative	You are in the informative cue condition. Red cues will appear over men who will approach a soldier, pull out a knife and attack.

Training Trials

The training trials commenced similar to Experiment 2 except with a VR specific eye-tracker as per the recommended Tobii procedure (*Calibration - Tobii Pro SDK documentation*, 2021). After the calibration step, the game scene was presented, showing the friendly soldier characters in a diamond formation around a central red square. The participant's eye-gaze location was visible within the VR scene via green dots. Experimenters instructed participants to check that the green dots aligned with their gaze when looking at the friendly soldiers. Experimenters also checked for alignment by observing the game scene on the external monitor. The eye-tracker calibration process could be repeated if the participant or experimenter determined the eye-gaze was inaccurate. The experimenter then informed the participant that a baseline SSVEP reading would be obtained via a calibration process. This calibration involved the participant watching a grey rectangle flashing at 15 Hz in the VR visual scene for one minute. The participant clicked on the red square when they were ready to commence the first trial.

The remainder of the trial proceeded in the same manner as in Experiment 2. Trials lasted approximately 90 seconds (M = 96.15sec, SD = 21.84sec), ending when participants had successfully clicked on between 21 and 29 frequent threats. The exact number was randomly selected for each trial. The participants hit and miss totals were displayed after the trial finished. A rest period was then offered. This trial format was repeated twice, starting at the eye-tracker calibration stage. These three trials comprised the training trial block.

Critical Trial

The critical trial occurred immediately after the third training trial. The participant received no indication that this trial was different to the previous trial. The only change from the training trial was that the rare threat character appeared randomly from the left or right 45 seconds after the trial started while carrying a large rifle. After onset, the rare threat traversed to the middle of the scene, stopped, faced the participant for two seconds and then proceeded to exit the scene opposite their onset location. The rifle disappeared if the participant clicked on the character, and then the character continued on their path unarmed, now appearing similar to a distractor character. They were visible for approximately 23 seconds (M = 23.28, SD = 3.08). Each trial ended when the rare threat exited the scene resulting in random trial lengths of approximately 81 seconds (M = 84.64sec, SD = 1.06sec).

Full-attention Trial

The full-attention trial was similar to the critical trial, except that participants were instructed not to search for and click on the frequent threats. The verbal instructions were that the system would automatically click any frequent threats attacking the friendly soldiers. Experimenters also reminded participants to continue looking for any threats. Each trial ended when the rare threat exited the scene resulting in random trial lengths of approximately 83 seconds (M = 83.01 sec, SD = 0.74sec).

Other Measures

Participants were provided the same video game experience, cognitive load and experiment naivety measures as Experiment 2. As in Experiment 2 only the naivety measures were analysed and reported in this thesis.

Measures

Behavioural Measures & Character Tracking. Unity logged all character's current world-coordinate locations and regions of interest (ROIs) during all trials at a 90Hz sample rate, matching the headset refresh rate. Additionally, click time & location were logged for any frequent or rare threat character clicks, forming the basis for these characters' response time and hit rate measures. Some gaze measures also used this data.

Because all characters were constantly moving throughout the scene, character ROIs were defined dynamically using Unity "world coordinates". These 3D ROIs were a cuboid 2m high, 1m wide, and 1m deep, centred on the midpoint of each character's skeletal frame (approximately their waist). The VR system provided depth perception by scaling objects based on size and distance from the headset's location in the Unity world-space coordinate system. Visual angles of each ROI varied between 2.8° and 9.9° (*M* = 6.6°).

Gaze measures. Unlike screen-based gaze samples in previous studies that are recorded as 2D co-ordinates, within VR, gaze samples are recorded as *gaze vectors*, at a sampling rate of 120 Hz. The nature of the visual scene prevented reliable calculation of a precise gaze end-point within the 3D scene, resulting in the processing of gaze vectors in a manner comparable to the ray method discussed in Weber et al. (2018).

Fixation calculations used a simple velocity threshold (Salvucci & Goldberg, 2000). Any two or more sequential gaze vectors with a velocity less than 100vd/sec (Salvucci & Goldberg, 2000) were combined and classified as a fixation. Velocity was determined by calculating the 3D angular difference between the current gaze-vector and the previous gaze-vector and then dividing this by the time difference in seconds between the two samples. Fixation vectors were recorded by using the means of the origins and unit-vectors from the gaze vectors allocated to the fixation. As per Experiment 2, trial durations differed slightly between participants due to characters' non-deterministic paths. Thus, the number of fixations was divided by trial duration (in seconds) in order to define the fixations per second metric.

Fixations were assigned to characters when a ray based on the fixations vector intersected with the characters' cuboid-shaped ROI for the samples in which the fixation occurred. Noting that a single fixation could be assigned to two or more ROIs if the characters visually overlaid each other at that moment. 33% of fixations during critical trials were not assigned to a character, 38% were assigned to only one character, and 29% were assigned to two or more characters. This list of ROI-classified fixations was used to define character-based metrics for each character type (rare threat, frequent threat, distractors, friendly soldiers). The character-based metrics were proportion dwell time for all characters, along with time to first fixation for the rare threat.

Dwell time proportion for a given character (e.g., distractor #2) was calculated as the total time spent fixating upon that character divided by the total time that the character was *potentially* visible in the scene. Within this VR study, characters were always nominally visible once they entered the scene, but the participant's head movements controlled if a character was within the participants’ field of view at any given moment. In order to not subject gaze measurement to an additional dependency (head movement), the standardizing denominator remained the time during which the characters were within the scene, and thus *potentially* viewable, rather than within the participants’ current field of view. In comparison, within Experiment 2, programming logic alone determined if a character was visible on screen at any moment. These per-character dwell time proportions were averaged across all characters of that type (e.g., all 21 distractor characters) to yield a mean dwell time proportion for that character type (e.g., distractors) per trial per person. For the critical rare threat character, the denominator was not the total time on-scene but the percentage of time they constituted a threat (i.e., before they were detected or left the scene), which was practically implemented as the period of time they were drawn carrying a gun.

The global search measure used to quantify the distribution of gaze points across the scene (distance to nearest friendly character) was conceptually the same as the Experiment 2 equivalent but with necessary adjustments for the Unity 3D coordinate system compared to 2D screen coordinates. Importantly, the unit of measure was no longer pixels but rather in-world meters.

Steady State Visually Evoked Potentials (SSVEP). VR raised practical considerations with EEG data collection as the combination of a VR headset and an EEG headset led to sub-optimal fitment of both devices. See Figure 27 above for an example of the headset strap interference with the EEG electrode locations. The bulk and weight of the VR headset meant there was no practical alternative to using the supplied mounting straps without the risk of the headset becoming uncomfortable or moving around. Therefore, some electrodes were not used when fitting the EEG headset due to the VR headset strap locations. Resulting in a restricted electrode montage (FP1, AF3, F7, F3, FC1, FC5, C3, P7, Pz, O1, Oz, O2, P8, C4, FC6, FC2, F4, F8, AF4, FP2, Mast1, Mast2, HEOGleft, HEOGright, VEOGlow, bpHEOG, bpVEOG, ERG1, ERG2).

During participant setup experimenters were requested to monitor any VR headset interference with the occipital lobe electrode locations O1, Oz, and O2, which are commonly used to record SSVEP signals (Norcia et al., 2015). Experimenter feedback indicated no interference between the VR headset and these locations. All remaining EEG/SSVEP equipment, setup, and data processing were as per Experiment 2.

Results

Behavioural Results

Rare Threat

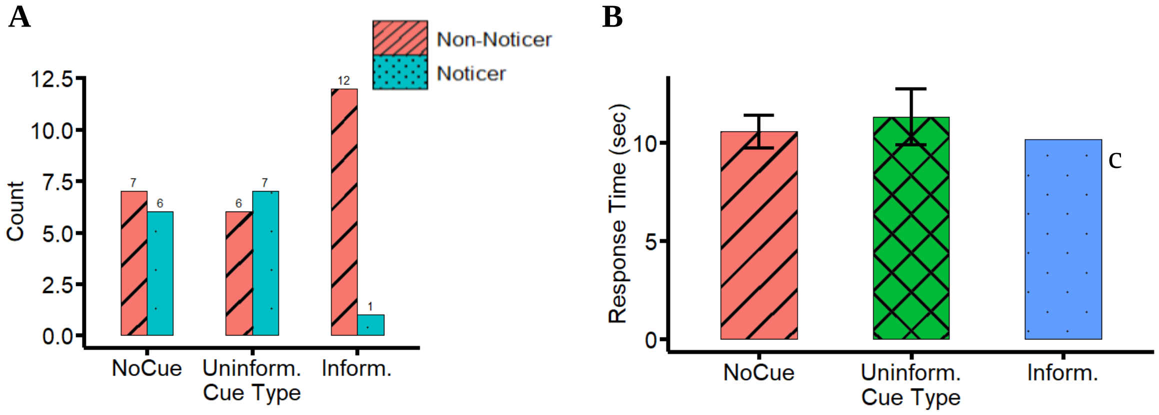
Behavioural responses to rare threat characters were the primary measure of inattentional blindness. Participants were classified as a noticer if they clicked on the rare threat character in the critical trial and a non-noticer if they failed to click. Across all conditions 64.1% (n = 25) of participants failed to click on the rare threat in the critical trial and were classified as non-noticers and therefore considered to have experienced inattentional blindness.

To investigate the effect of cueing on noticing, analysis was performed comparing the no-cue, uninformative and informative cue conditions during critical trials. A G-test revealed a significant difference between the proportion of non-noticers and noticers between the three cue type conditions, $G(2) = 7.98$, $p = .019$, $V = 0.42$, $BF_{10} = 20.02$ (Figure 29A).

A oneway ANOVA found no significant effect of cue type on rare threat response times during the critical trials, $F(2,11) = 0.12$, $p = .885$, $\eta^2_G = 0.02$, $BF_{10} = 0.40$ (Figure 29B). However, this analysis was compromised by the lack of responders (noticers) in the informative cueing condition (n = 1), which meant that very little response time data was provided from this group.

Figure 29

Rare Threat Behavioural Responses by Cue Type During Critical Trials.



Note. (A) Noticers vs non-noticers. (B) Response times in seconds. Error bars represent standard error. (C) Informative cue condition represents a single noticer, hence no error bar.

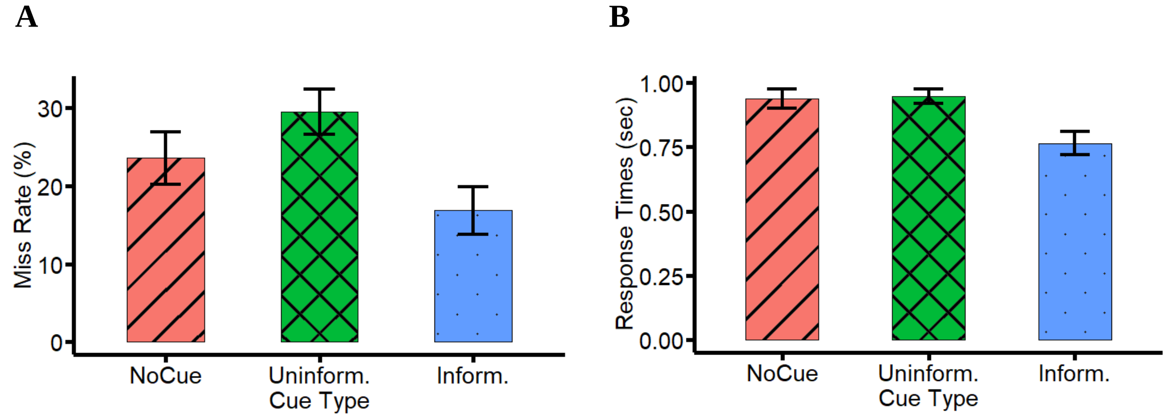
These outcomes combined with visual inspection of Figure 29 suggest that an uninformative cue has limited effect on noticing rates or response times. However, providing an informative cue results in significantly higher inattentional blindness.

Frequent Threats

Analyses of the effect of how informative a cue was (no-cue, uninformative, informative) on the behavioural responses towards the frequent threat characters during critical trials were conducted. A oneway ANOVA showed a significant effect of cue type condition on miss rates, with anecdotal evidence, $F(2,36) = 4.16$, $p = .024$, $\eta^2_G = 0.19$, $BF_{10} = 19.23$ (Figure 30A). Posthoc pairwise t-test results (Table 16) show that frequent threat miss rates are significantly lower in the informative cue condition than in the no-cue or uninformative cue conditions.

Figure 30

Frequent Threat Behavioural Responses for All Conditions During Critical Trials.



Note. (A) Miss rate (%). (B) Response times in seconds. Error bars represent standard error.

Table 16

Pairwise T-Test Results for Frequent Threat Miss Rates for All Cue Type Conditions During Critical Trials.

Trial type 1	**Trial type 2**	***t(df)***	***p-adjust***	BF_{10}
No-cue	Uninformative	1.35(11)	.559	0.69
No-cue	Informative	1.54(11)	.400	0.81
Uninformative	Informative	2.88(11)	.020	7.49

Note. p-adjust based on Bonferroni adjustments.

A oneway ANOVA showed a significant effect of cue type condition on response times, with strong evidence, $F(2,36) = 7.39$, $p = .002$, $\eta^2_G = 0.29$, $BF_{10} = 19.23$ (Figure 30B). Posthoc pairwise t-test results (Table 17) show that frequent threat response times were significantly higher in both the no-cue and uninformative cue conditions than in the informative cue condition but not significantly different when comparing the no-cue and uninformative conditions.

Table 17

Pairwise T-Test Results for Frequent Threat Response Times for All Cue Type Conditions During Critical Trials.

Trial type 1	**Trial type 2**	***t(df)***	***p-adjust***	BF_{10}
No-cue	Uninformative	0.17(11)	1.000	0.37
No-cue	Informative	3.24(11)	.008	6.80
Uninformative	Informative	3.41(11)	.005	15.23

Note. p-adjust based on Bonferroni adjustments.

In summary, these results reveal that providing an uninformative cue leads to similar behavioural responses to the frequent threat as providing no cue at all. Whilst provision of an informative cue on all frequent threats improves behavioural responses.

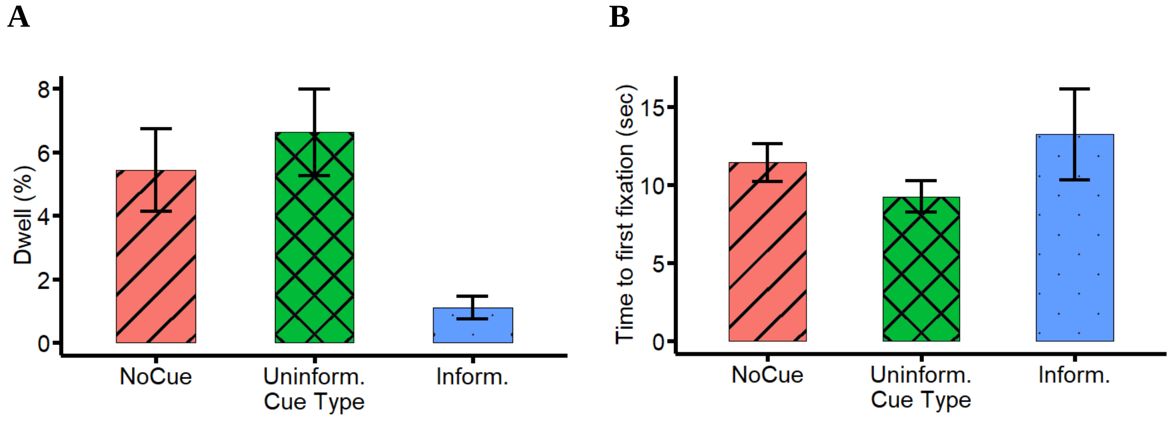
Eye Tracking

Rare Threat

The effect of cueing on rare threat dwell during critical trials was analysed using a oneway ANOVA, with three levels: informative, uninformative and no cue. This analysis found an effect of cue type condition on dwell towards the rare threat character during critical trials, with anecdotal evidence, $F(2,20) = 3.60$, $p = .046$, $\eta^2_G = 0.27$, $BF_{10} = 1.82$ (Figure 31A), excludes 16 participants who did not fixate on the rare threat character. Posthoc pairwise t-test results (Table 18) show that frequent threat response times are significantly higher in the uninformative cue condition than in the informative cue condition but not significantly different when comparing the no-cue to either the informative or uninformative conditions.

Figure 31

Eye Gaze Towards Rare Threat Characters Comparing all Conditions During Critical Trials.



Note. (A) Dwell as a proportion of time gun was visible. (B) Time to first fixation in seconds. Error bars represent standard error.

Table 18

Pairwise T-Test Results for Rare Threat Dwell for All Cue Type Conditions During Critical Trials.

Trial type 1	**Trial type 2**	***t(df)***	***p-adjust***	***BF₁₀***
No-cue	Uninformative	0.61(16)	1.000	0.47
No-cue	Informative	1.85(9)	.239	3.78
Uninformative	Informative	2.67(15)	.044	2.89

Note. (A) p-adjust based on Bonferroni adjustments. (B) Degrees of freedom varies irregularly within this table due to not all participants in all groups dwelling on the rare threat character, resulting in mismatched group sizes.

A oneway ANOVA found no effect of cue type condition on time to first fixation towards the rare threat character during critical trials, $F(2,20) = 1.75$, $p = .199$, $\eta^2_G = 0.15$, $BF_{10} = 0.74$ (Figure 31B). Note that this analysis was challenged by the observation that 16 participants did not register a single fixation on the rare threat character during the critical trial.

These eye gaze outcomes indicate that those in the informative cue condition gaze for significantly less time on the rare threat characters than those in the other two conditions.

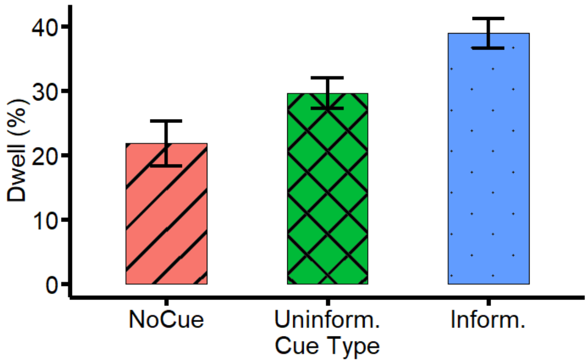
However, there was no significant differences in how fast participants where to first fixate on the rare threat after onset.

Frequent Threat

The effect of how informative a cue was on frequent threat dwell during critical trials was analysed. A oneway ANOVA revealed an effect of cue type condition on frequent threat dwell during critical trials, with very strong evidence, $F(2,36) = 9.50$, $p < .001$, $\eta^2_G = 0.35$, $BF_{10} = 63.51$ (Figure 32). Posthoc pairwise t-test results (Table 19) show that the informative cue trials had a significantly higher dwell than the no-cue but there was no difference between the uninformative cue condition and the no cue and informative cue conditions.

Figure 32

Dwell Proportion for Frequent Threat Characters During Critical Trials.



Note. Error bars represent standard error.

Table 19

Pairwise T-Test Results for Frequent Threat Dwell, All Conditions During Critical Trials.

Trial type 1	**Trial type 2**	***t(df)***	***p-adjust***	BF_{10}
No-cue	Uninformative	1.99(11)	.163	1.24
No-cue	Informative	4.35(11)	< .001	61.10
Uninformative	Informative	2.36(11)	.071	5.13

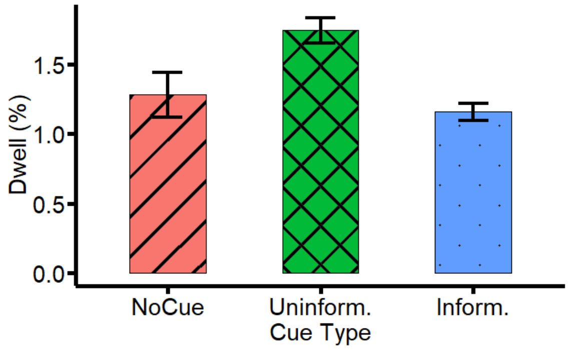
Note. p-adjust based on Bonferroni adjustments.

Distractors

The effect of cueing on distractor character dwell during critical trials was analysed. A oneway ANOVA revealed an effect of cue type condition on distractor character dwell during critical trials, with very strong evidence, $F(2,36) = 7.29$, $p = .002$, $\eta^2_G = 0.29$, $BF_{10} = 18.20$ (Figure 33). Posthoc pairwise t-test results (Table 20) show that the uninformative cue condition had a significantly higher dwell than the uninformative cue and no-cue conditions but there is no difference between the no-cue condition and informative cue condition.

Figure 33

Dwell Proportion for Distractor Characters During Critical Trials.



Note. Error bars represent standard error.

Table 20

Pairwise T-Test Results for Distractor Character Dwell, All Conditions During the Critical Trials.

Trial type 1	Trial type 2	*t(df)*	*p-adjust*	BF_{10}
No-cue	Uninformative	2.86(11)	0.021	2.96
No-cue	Informative	0.76(11)	1.000	0.44
Uninformative	Informative	3.62(11)	0.003	> 100

Note. p-adjust based on Bonferroni adjustments.

Cued Characters

In addition to asking whether participants biased their gaze towards frequent threats, we considered whether cueing condition (informative or uninformative) influenced gaze bias towards cued characters. An independent samples t-test revealed greater dwell towards cued characters in the informative ($M = 5.61$, $SD = 1.59$) than the uninformative ($M = 4.14$, $SD = 1.84$) cue conditions during the critical trials, $t(24) = 2.19$, $p = .049$, $d = 0.86$, $BF_{10} = 1.94$.

Noticers

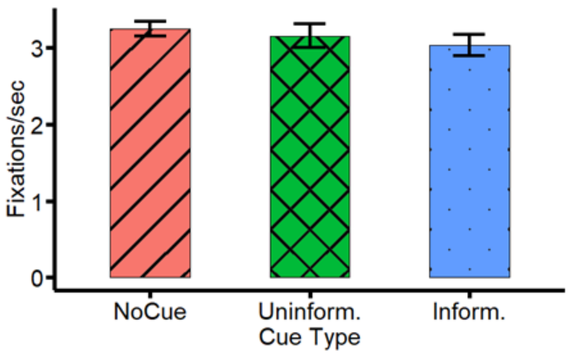
Analyses were planned to investigate changes in dwell towards cued and uncued characters for those classified as noticers and non-noticers. Such analyses might allow conclusions regarding whether time spent looking at the cues led to, or at least was associated with, failing to notice the rare threat. However, no results will be presented as it was deemed an inappropriate comparison due to the extreme noticing rate differences between the uninformative (54%) and informative (8%) cue condition. With only a single noticer in the informative cue condition the outcome would be skewed and not a true comparison between cued/not-cued characters by noticer status.

Global Search Behaviour

Within Experiment 2, participants global search behaviour differed between conditions. Therefore, the similar measures were analysed for the present experiment (fixations per second and distance to nearest friendly soldier). A oneway ANOVA revealed no effect of cue condition on fixations/sec (Figure 34), $F(2,36) = 0.70$, $p = .505$, $\eta^2_G = 0.04$, $BF_{10} = 0.29$.

Figure 34

Fixations/sec for All Conditions.

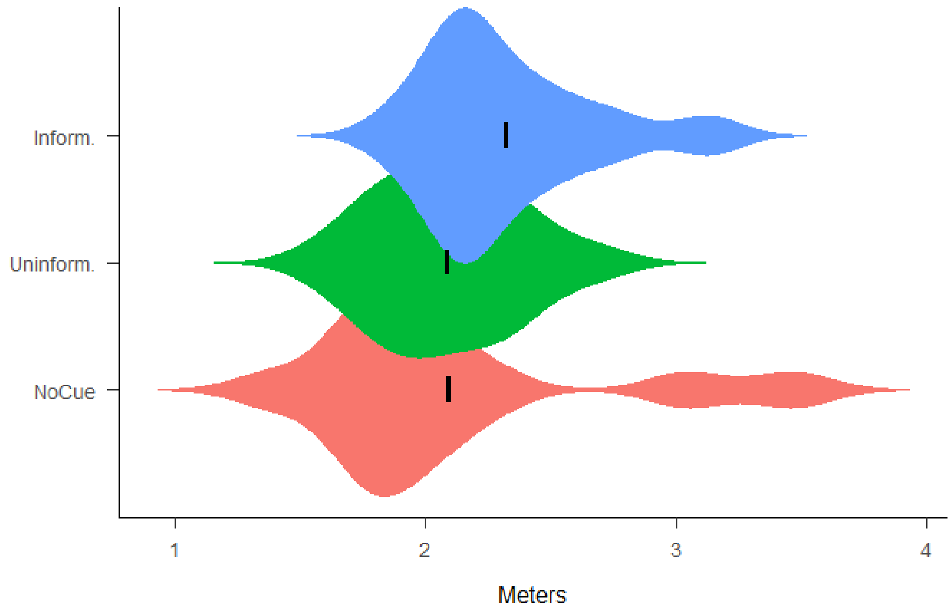


Note. Error bars represent standard error.

Search breadth was measured via the distance to nearest friendly soldier metric. Although the data appeared to show the same numeric trend as was evident previously, a oneway ANOVA revealed no effect of cue type condition on distance to nearest friendly soldier, $F(2, 35) = 1.26$, $p = .295$, $\eta^2_G = 0.07$, $BF_{10} = 0.43$ (Figure 35).

Figure 35

Violin Plot of Distance in Meters of Fixations From Nearest Friendly Soldier Character by Cue Type.



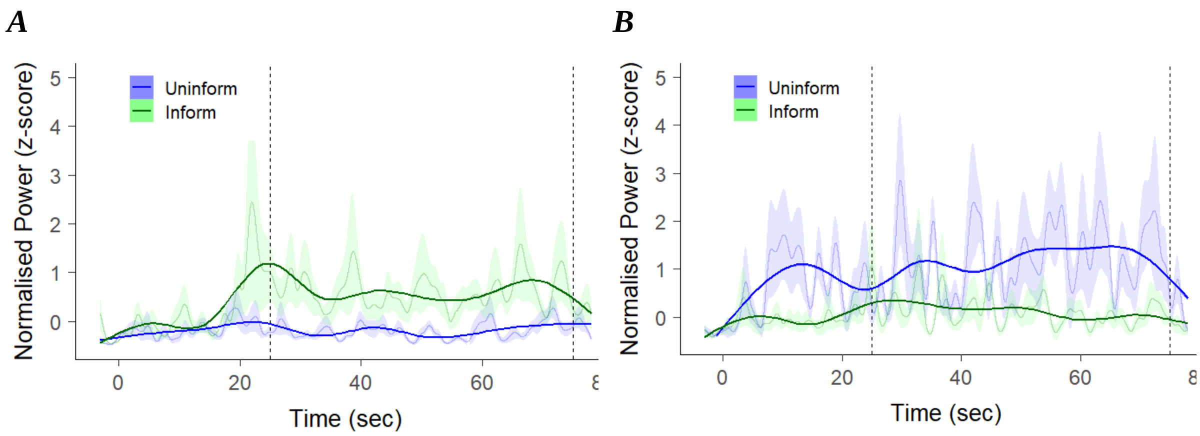
Note. Bars show mean distance in meters.

SSVEP

Exploratory attention analysis was performed using SSVEP, which provided a combined measure of overt and covert attention. For comparisons, the SSVEP signal for the informative and uninformative cue conditions was normalised against the signal from the no-cue condition, resulting in a time series of z-scores (Figure 36). The initial review of the results suggested that the SSVEP signal was too weak to be informative. In the critical trials, the SSVEP power in the uninformative cue bordered on zero, indicating it was roughly equivalent to the no-cue trial. Similarly, in the full-attention trial, the informative cue condition was approximately equal to the no-cue trial. These power levels are unlikely because a portion of the SSVEP response is obligatory; it should be generated with some power whenever flickering is present in the visual field, even if the participant is not actively attending to that stimulus (Norcia et al., 2015). As there were multiple flickering stimuli present throughout the scene for both informative and uninformative conditions, an observation of no elevation in SSVEP power relative to baseline is more likely to be the product of experimental or measurement insensitivity (in the context of large head and eye movements) than genuinely reflecting the absence of an SSVEP response. Even occasional overt attention towards a cued character would be expected to increase SSVEP power beyond these levels. The most likely explanation is interference from the VR headset and increased muscle artifacts in the EEG (due to the unconstrained head movements) resulted in most of the EEG signal being eliminated by the cleansing/filtering stages of the SSVEP processing pipeline. Due to low confidence in the SSVEP signal the decision was made not to interpret further the SSVEP results. These concerns will be discussed in more detail in the General Discussion chapter.

Figure 36

SSVEP Normalised Power Comparison Between Informative and Uninformative Cue Conditions



Note. (A) Critical trials. (B) Full-attention trials. Bold lines represent smoothed results with raw values underneath. Vertical dotted lines indicate the range of values used for inferential tests. See the measures section above for further details. Error bars represent standard error.

Full-Attention Trials

The full-attention trials provided an opportunity to investigate if the rare threat character was plainly visible when participants where unburdened by the demanding frequent threat search task. Expectancy of the rare threat was also expected to be higher in the full-attention trials due to a combination of task-instructions and potentially having noticed the rare threat during the critical trial. Full-attention trial rare threat character data was grouped into a 2 (hits, misses) x 3 (no-cue, uninformative, informative) table (Table 21). As only a single participant failed to notice the rare threat character no inferential analysis was possible or required.

Table 21

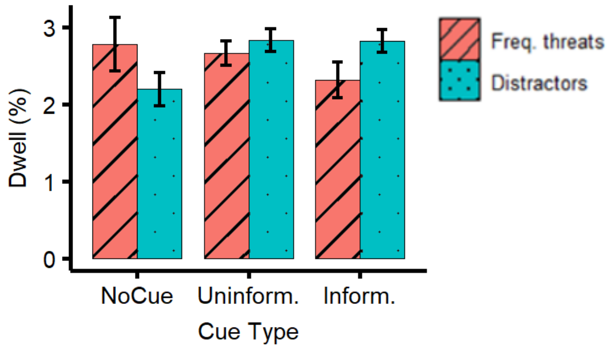
Contingency Table Showing Rare Threat Character Hits and Misses for All Cue Type Conditions During Full-Attention Trials.

	No-cue	Uninformative	Informative	Total
Hits	12	13	13	38
Misses	1	0	0	1
Total	13	13	13	39

A mixed ANOVA was performed to evaluate the effect of cue type condition type (no-cue, uninformative, informative) and character type (distractor, frequent threat) on dwell (Figure 37) during the full-attention trials. The within subjects main effect for character type was not significant, $F(1, 35) = 0.04$, $p = .850$, $\eta^2_G < 0.001$, $BF_{10} = 0.24$. The between subjects main effect of cue type on dwell was not significant, $F(2, 35) = 0.74$, $p = .485$, $\eta^2_G = 0.02$, $BF_{10} = 0.24$. The interaction between character and cue type conditions on dwell was significant, $F(2, 35) = 3.59$, $p = .038$, $\eta^2_G = 0.09$, $BF_{10} = 0.06$. Due to the very small effect size and strong evidence against an effect, the interaction was determined to be uninteresting.

Figure 37

Dwell by Cue Type for Frequent Threat and All Distractor Characters During Full-Attention Trials.



Note. Distractors include all distractor types. Error bars represent standard error.

These results indicate that during the full-attention trials when participants are unburdened by the primary search task the rare threat character was plainly visible and that any effects of cue type on frequent threats or distractor characters observed in the critical trials was no longer present.

Inattentional Blindness Awareness

Awareness of inattentional blindness and the gorilla experiment (Simons & Chabris, 1999) were queried post-experiment; 26% of participants reported awareness of inattentional blindness, and 44% were aware of the gorilla experiment. No participants indicated associating this experiment with inattentional blindness or the gorilla experiment.

Discussion

The present experiment replicated the previous experiment (Experiment 2) within a VR environment to continue investigating the potential for AR cues to induce inattentional blindness. A change anticipated to provide participants with a more spatially realistic visual experience (Brookes et al., 2020; Kisker et al., 2019) and lead to participants' search strategies more closely aligning with how they would search similar contexts in the real world. With the closer mapping of the projected visual scene to perceived visual angles, there may have been less opportunity for visual clutter and consequently more opportunity to observe an effect of attentional tunnelling. In fact, the findings in the VR task generally fitted the predictions of an attentional tunnelling account. Informative cues were particularly helpful in detecting frequent threats, and appeared to bias attention to cued characters, but also markedly reduced sensitivity to an uncued and unexpected additional threat. Each finding is considered in turn.

Informative cues helped the detection of frequent threats. Miss rates towards the frequent threat characters improved when an informative cue was provided compared to providing an uninformative cue. An informative cue also produced faster response times to frequent threats than an uninformative or no-cue. Eye-tracking measures show that compared to the other conditions, those in the informative cue condition increased dwell towards the frequent threat characters (the only cued characters) and reduced dwell towards the rare threat and distractor characters. This combination of evidence demonstrates that those in the informative cue condition remained focused on the cued frequent threat characters to the detriment of the search for uncued threats during the critical trials. Participants in the cued conditions responded differently to frequent threats based on the type of cue provided, suggesting the cue's utility, and not its salience, positively impacted performance. Results that align with an attentional tunnelling account of inattentional blindness due to the cues facilitating the search for cued threats (Wickens & Alexander, 2009; Yeh & Wickens, 2001).

Inattentional blindness rates were globally high. Overall, the majority (64%) of participants were classified as non-noticers, i.e., failed to notice the rare threat character during the critical trial. The noticing rates for the no-cue and uninformative cue conditions were comparable, with ~50% in each condition classified as non-noticers. However, almost perfect inattentional blindness was observed in the informative cue condition: only one participant out of thirteen clicked on the rare threat character. Further, this participant volunteered after the experiment that they had accidentally clicked on the rare threat in the course of trying to target a frequent threat character, and only became aware of the rare threat character through that error (they were nonetheless scored as a noticer). This constitutes further evidence that participants' behavioural responses varied depending on the type of cue provided rather than on the mere presence of cues (as seen in Experiment 2).

Despite the level of inattentional blindness in the informative cue condition being extremely high (92%), it is still at a level supported by prior evidence. A similarly high rate of ~84% was observed in the informative cue condition within Experiment 2. Additionally, in a simulated AR-supported surgical procedure, Dixon et al. (2013) found blindness rates > 90%, unlike non-AR-supported participants who experienced a much lower rate (59%).

Because this was our first VR study, the high blindness rate might have been due not to inattentional blindness but instead a product of rendering the rare threat character too subtly to be reliably detected in the VR-projected scene. However, the 100% noticing rate for the same participants in the full-attention trial and the ~50% noticing rate in the other conditions during the critical trial strongly suggest that the rare threat character was adequately visible. Both under normal viewing conditions (as evident in the full-attention trials) and when salient cues were present onscreen (as evident in the uninformative cueing condition on both trials). Instead, high rates of inattentional blindness appear to be an effect of providing informative cues. Within the inattentional blindness literature this contrast between critical and full-attention trial noticing rates is commonly considered a valid indicator of how obvious or salient the unexpected object is (Mack & Rock, 1998).

One of the motivators for this VR experiment was investigating the effect a more spatially realistic scene might have on inattentional blindness outcomes. Of specific interest was the suspicion that in Experiment 2, visual clutter increased inattentional blindness in the uninformative cue condition. Prior studies such as Schöne et al. (2020) demonstrated reduced inattentional blindness in VR compared to screen-based equivalents, providing support for this suspicion. However, this did not occur within our VR experiment, as those in the uninformative cue condition showed similar behavioural responses to the no-cue condition for both frequent and rare threat characters. The uninformative cue did not appear to significantly inhibit the detection of the frequent threat characters (contrary to Experiment 2) nor reduce the detection of the uncued, rare threat character. Additionally, eye-tracking indicated that dwell patterns across character types in the uninformative condition were similar to those in the no-cue condition. The most parsimonious explanation of this data set is that within the uninformative cue condition in the present experiment the uninformative cues were successfully ignored, unlike in Experiment 2. Suggesting a reduced role for visual clutter in the present experiment.

Effect of Cues on Global Search

Participants showed no significant differences between conditions in the global gaze measures. This finding was unsurprising for the fixations per second measure as this replicated the Experiment 2 outcome. However, it was surprising that the search breadth metric (distance to nearest friendly soldier) showed non-significant results, contrary to the pattern seen in Experiment 2. Comparison of the distance to nearest friendly soldier plot

(Figure 35) with the equivalent from Experiment 2 (Figure 24) shows similar numeric patterns, even though inferential analyses do not support this interpretation.

The lack of an effect on search breadth is likely due to a design change where the world depicted in the VR game scene was markedly narrower than the Experiment 2 equivalent, due to the use of obscuring black “curtains” (see Figure 38). Although the width of participants’ field of view was retained across experiments, there was a sizable reduction (~33%) in the width of the scene potentially depicted (here “potential” is used in the same manner as described in the gaze analysis calculations). This change likely impacted upon the breadth metric. Participants had less scene space in which to spread their search beyond the friendly soldier locations, thereby reducing the sensitivity of the breadth measure.

This scene design change was made as part of the migration to the VR environment. The black “curtains” that narrow the scene width were included to prevent participants from observing certain video game mechanics. For example, the curtains provided an invisible space allowing characters to onset, offset, and reverse directions naturally (actions that previously took place offscreen). Perhaps more importantly, it also stopped participants from orienting their field of view away from the core scene; an action that was impossible in the screen-based task. Interestingly, this change would tend to increase the proportion of the display occupied by the rare threat, yet a net decrease in detections of the rare threat was observed. This observation is in direct contradiction to the explanation Schöne et al. (2020) offered for their finding of lowered inattentional blindness in the VR format relative to the screen-based format. Future studies based on the present paradigm should consider how important search breadth measures are to the research question and appropriately adapt both the scene and measure designs.

Figure 38

Comparison of Experiment 2 and 3 Game Scenes.

A ***B***

Note. (A) Experiment 2 critical trial scene as presented on a computer monitor to participants. (B) Experiment 3 critical trial screen as viewed on an external monitor, with participants seeing the same scene via VR headset. Contrast and scaling modified for publication purposes.

Conclusion

The present study shows that providing task-relevant AR cues in the informative cue condition increased search performance for cued targets at the expense of searching for uncued targets, replicating similar results from Experiment 2. However, informative cueing led to a clear performance cost in the present experiment. Task-relevant cues attracted and held attention, thereby preventing the allocation of attentional resources to uncued stimuli. In contrast, participants who received an uninformative cue performed comparably to those given no cueing, suggesting they were able to ignore uninformative cues in the present experiment, unlike in Experiment 2. Cumulatively, these findings suggest that any visual clutter or distraction generated by the onscreen cues exerted less of an impact upon search in the present experiment, and instead the elevated inattentional blindness rate was more likely the product of attentional tunnelling.

Chapter 5: General Discussion

The overarching aim of this thesis was to investigate if providing AR visual cues during an operational search task would increase inattentional blindness to uncued stimuli. Additionally, within operational search contexts, noticing rates of a single target type is rarely the only measure of interest. Returning to the motivating example used throughout this thesis, a soldier on overwatch duty must not just notice a single type of threat but *all* threats rapidly enough to respond promptly. Failing to react fast enough can result in outcomes equal to those possible for a failure to notice. Hence, this thesis focuses on measuring AR visual cues' impacts beyond facilitation of search for AR-cued objects, and indeed beyond any single measure of performance (i.e., noticing rate). This chapter summarises the main findings of the present research, discusses these findings in relation to key hypotheses and concludes by considering the implications for AR in operational search and future research.

Summary

Experiment 1 investigated AR visual cues using a within-subjects design. Participants received both cued and uncued trials, with the rare uncued threat appearing three times in each critical and full-attention trial. Results showed the expected improvement in behavioural responses to the cued frequent threat characters and no significant cost in behavioural responses to the uncued rare threat character. Despite the lack of costs observed, there was evidence that AR cues had an effect, with increased attention towards cued characters and a broader pattern of eye gaze across the scene during cued trials. The repeated onsets of the uncued rare threat character (three per trial) also highlighted the importance of cue expectancy. Participants were far more likely to continue noticing the rare threat character after the first time they noticed it.

Experiment 2 extended Experiment 1; by moving to a between-subjects design, such that the uncued rare threat appeared only once each in the critical and full-attention trials.

Experiment 2, also introduced a new control condition: the uninformative cue condition. The increased power of the between-subjects design revealed a cost to providing AR visual cues. Inattentional blindness rates were higher in the two cued conditions (informative, uninformative cues) than in the uncued condition. The presence of uninformative cues impaired detection rates for the frequent threats (relative to the no cue condition), suggesting some reliance on the unhelpful cues. Yet the uninformative cue condition's eye gaze and attentional patterns were more similar to the uncued than the informative cue condition. Overall, the way participants interacted with the cues differed between the informative and uninformative conditions but both led to inattentional blindness.

Experiment 3 moved the inattentional blindness protocol into a VR environment. The only changes were those necessary for the experiment to work in VR (see Chapter 4). The informative cue condition results broadly reflected those seen in the screen-based Experiment 2. However, for the uninformative cue condition, most behavioural, eye gaze, and attentional allocation measures for the uninformative cue condition were now most similar to the uncued condition. In summary, informative cues retained their influence on attention and responding within VR, but the uninformative cues exerted a weaker effect on overall search performance in the VR domain (Experiment 3) than in the screen-based task (Experiment 2).

Overarching Hypotheses

Three overarching hypotheses were addressed by the three experiments: (1) AR visual cues will reduce cognitive load, therefore reducing inattentional blindness; (2) AR visual cues will increase visual clutter of the scene, making it more difficult to search; (3) The visual salience and task-relevance of the cues will induce attentional tunnelling to the cues. In addition, there was likely a fourth important contributor to the pattern of findings. Expectancy, specifically, expecting more than one threat type – likely impacted on noticing rates in all experiments, despite attempts to minimise and control for this factor. The cumulative data (behavioural, gaze, EEG) across all three experiments were jointly used to evaluate each hypothesis.

Cognitive Load

The cognitive load hypothesis represents what might be expected to occur if the marketing claims of AR supports are to be believed on their face. That is, providing visual support to participants should help their overall search performance. The hypothesis is formed from two elements. The first element is the assumptions that providing salient and informative AR visual cues will facilitate the search for cued objects, making that search task easier (Eisma et al., 2020; Hughes-Hallett et al., 2015). This idea is generally well supported in the basic science (Posner, 1980) and applied literature (e.g., Dixon et al., 2013; Yeh & Wickens, 2001). This present thesis repeatedly found evidence to further support this claim. Whenever an *informative* cue was provided for frequent threat characters, hit rates and response times improved in parallel with increased attention to cued characters.

The more interesting element in the cognitive load hypothesis is the second claim. Within the inattentional blindness literature there is evidence that reducing cognitive load during a task significantly reduces inattentional blindness for visual stimuli shown unexpectedly during that task (e.g., Mack & Rock, 1998; Simons & Chabris, 1999).

While several studies have examined this hypothesis either directly, or indirectly, and the global findings are generally mixed (Hutchinson, 2019). Those studies that most directly tested the role of load (e.g., Chabris et al., 2011; Mack & Rock, 1998) found evidence that inattentional blindness increases under load. Yet it is important to note that the relationship between cognitive load and the specific parameters of the search task are not always clear. For example, consider the influence of the number of targets in the primary search task. In Mack and Rock’s classic inattentional blindness task, participants are asked to monitor a single stimulus for fine, rapid changes in size. In the Most et al. (2001) bouncing shapes task, people are asked to monitor several bouncing shapes. Was cognitive load higher to monitor one shape for small changes, or to monitor the gross positions of several? The answer remains unclear, particularly given the existing disagreements in how many discrete components of work constitute cognitive load (Kalyuga, 2011). What remains clear is that irrespective of potential differences in cognitive load induced by the primary search task in the various inattentional blindness paradigms, robust inattentional blindness was present in all instances.

In the present experimental series, the cognitive load hypothesis would anticipate that facilitating the search for frequent threats using informative cues would reduce cognitive load. In turn, the reduced cognitive load in the presence of informative cues should, reduce inattentional blindness. However, in no experiment did facilitating the search for cued characters improve inattentional blindness. The opposite occurred in Experiments 2 and 3; providing an informative cue increased inattentional blindness. It is, therefore, unlikely that the cognitive load hypothesis offers a valid explanation for the results found in the present research.

In potential support for the cognitive load hypothesis is one interesting but underpowered finding: Even though inattentional blindness was high in the informative cue conditions, those who did notice appeared to notice faster than those in the other conditions. A reduction in cognitive load may have facilitated a more comprehensive search strategy for some participants. What is unclear is if this was an anomaly due to low noticer numbers or an indication that those who did notice were potentially using a different search strategy. More specifically, did they accidentally notice or did the cues improve the search for a few participants? Future research that reduces the task's difficulty would increase the number of noticers and provide enough statistical power to investigate if this was a valid effect or an anomaly.

Attentional Tunnelling

Like the cognitive load hypothesis above, the attentional tunnelling hypothesis also starts by suggesting that cues facilitate the search for cued targets. Unlike the cognitive load hypothesis, the attentional tunnelling hypothesis asserts that the improved performance for the cued targets comes at the cost of uncued targets. Evidence for this hypothesis has been observed in studies investigating AR in domains such as surgery (Dixon et al., 2013; Hughes-Hallett et al., 2015), air-traffic control (Eisma et al., 2020), and aerial search (Yeh & Wickens, 2001). Specifically, valid and salient cues capture spatial attention, which results in lowered sensitivity to uncued or unexpected targets that lie outside of the cued set (Anderson et al., 2011; Theeuwes, 2010).

Attentional tunnelling has strong support from the informative cue condition results in the Experiment 2 and 3 critical trials. In those two critical trials, behavioural responses towards the cued frequent threat characters improved markedly. In addition, increased overt attention (gaze) was observed towards the validly cued characters, compared to both the uninformative and no cue conditions. Despite not reaching statistical significance, SSVEP results align with the gaze data, showing a tendency for informative cues to attract more attention than uninformative cues. Critically, these indicators of facilitated attention towards cued targets coincided with elevated rates of inattentional blindness in this group. The implied bias in attention to cued targets and away from uncued, unexpected targets strongly aligns with the predictions of the attentional tunnelling account.

Whilst attentional tunnelling likely explains the *informative* cue results, it does not provide a full account of the *uninformative* cue conditions across Experiments 2 and 3. On the attentional tunnelling account, the uninformative cues should not preferentially capture attention, by virtue of their inability to facilitate detection of the frequent threat events. These cues were 33% reliable in cueing frequent threats and research has regularly indicated that cueing systems less than approximately 70% reliable are unlikely to be used (Dixon et al., 2007), despite some evidence that as low as 60% are used in some contexts (Wickens et al., 2006). In essence, this account predicts that uninformatively cued events should be largely ignored, in a similar manner to other objectively salient but task-irrelevant aspects of the scene, such as the motion of distractor characters or visually bright sand-coloured ground texture.

Consequently, on this account, detection of cued targets should not be facilitated (nor impaired) relative to the no cue condition, and equally the uninformative cue should not reduce or increase noticing rates relative to the no cue condition. This was the pattern observed in Experiment 3 (in the VR format) but was not the pattern observed in Experiment 2. In Experiment 2, uninformative cues reduced detection accuracy for the frequent threats and also resulted in high rates of inattentional blindness, akin to that seen in the informative cue condition.

Visual Clutter

A visual clutter account suggests perceptual rather than cognitive mechanisms moderate search performance. Specifically, it is anticipated that visually salient cues in the scene increase visual complexity, which in turn slows and challenges serial search through the scene. Henderson et al. (2009) proposes that increases in visual clutter influence visual search outcomes similarly to increases in set sizes. Because the game scene is inherently dynamic and complex, the provision of nine large, bright, moving and flickering visual cues could be assumed to increase visual complexity and thus visual clutter. This hypothesis was primarily explored in Experiments 2 and 3, because the crucial test of this idea lies in the performance of participants when given an uninformative cue. No uninformative cue condition was used in Experiment 1.

The visual clutter account anticipates that merely adding the visual cues to the scene will hinder search by distracting gaze towards cued objects and away from the intended, endogenous search target. Yeh et al. (2003) whilst investigating providing information via helmet-mounted displays found that displaying task-irrelevant information increased clutter of the scene. Resulting in increased workload of the participants. Results such as this provide support for the visual clutter account.

What is unclear is to what extent the specific context of the scene and the type of visual clutter provided influences how the visual search is impaired. It could be argued that providing an uninformative cue is driving top-down impairment of search by increasing the number of stimuli that participants need to evaluate for task-relevance, similar to Yeh et al. (2003). Alternatively, it could be the influence of both bottom-up (e.g., visual salience) and top-down factors that is attracting attention as proposed by search models such as Wolfe (2007).

Crucially, a pure visual clutter account does not distinguish between uninformative and informative cues; the provision of cues of any kind should induce clutter and impair serial search. However, within the present experiments because the cued objects were themselves the search targets in the informative cue condition, the informatively cued condition is not particularly informative as to the visual clutter hypothesis. Instead, the visual clutter hypothesis is primarily evaluated in participants' responses to the frequent threat characters in the uninformative condition, and in their noticing rates for the rare threat. In Experiment 2, frequent threat characters were less well detected in the uninformative cue condition than the no cue condition. Further, rates of noticing the rare threat were low in this group. Together, these two findings suggest that visual clutter offers a good account of the uninformative condition in Experiment 2. The mere presence of the uninformative cues reduced overall search performance.

By contrast, in Experiment 3, the uninformative cue condition demonstrated little evidence in support of the visual clutter hypothesis. Participants largely appeared to have ignored the uninformative cues in the VR format. Uninformative cues did not impair (nor facilitate) search for the frequent threat characters, and detection rates for the rare threat were comparable to the no cue condition (and much higher than the informatively cued group). The informative cue group performed quite differently to the uninformative cue group, demonstrating that cue validity affected performance more than the mere presence of cues. This pattern of findings did not support the visual clutter hypothesis.

Beyond the Experiment 3 data, the visual clutter account is also challenged by the full-attention control conditions present in each experiment. Specifically, the full attention control condition presented participants with the identical scene to the critical trial, but participants were no longer required to search for the frequent (knife-wielding) threats. Search for the unexpected, uncued rare threat should have been equally impacted by the presence of visual cues in the critical trial as in the full-attention trial, and thus detection rates for the rare threat should be equal across trials under the visual clutter account alone. That is, the presence of bright moving targets should have hampered detection of the rare threat in both instances. However, this was not the case; detection of the rare threat in the full attention control trials was essentially at ceiling.

It is possible that because the full attention trial occurred after the critical trial, then improved detection on the second occurrence was merely the product of more exposure to the rare threat; approximately 40s (of 200s) relative to approximately 20s (of 100s). However, such an improvement due to exposure could be expected to be incremental. If 36% of participants noticed the rare threat the first time it appeared (as in Experiment 3), then we could expect 36% of the remaining, unaware participants to notice the rare threat the second time it appeared. This would lead to detection rates of 59% by the second trial, which is much lower than the near perfect detection rates observed. In sum, visual clutter alone cannot account for the difference in performance between the critical and full attention trials, nor the performance of the uninformative condition in Experiment 3. Instead, at least one additional element is required to explain participants' performance. Notably, the critical trials and the full attention trials differed in participants' expectancy. The changed instructions for the full attention trials imply the presence of a threat or event that is distinct to the frequent threats, thereby manipulating participant expectancy.

Expectancy

Managing expectancy is central to all inattentional blindness paradigms (Mack & Rock, 1998), within the present research, expectancy was a critical recurring consideration. As a reminder, expectancy within the present context refers to a participant's expectations of the possibility of more than one threat type. Each experiment's trial sequences and instructions were created to manage participants' expectancy. From the perspective of our present research question, it was important to manage participants' expectations during the critical trial as the focus was on the impact of cueing on the detection of unexpected, *rare* threats.

To this end, the task instructions needed to be simultaneously clear but ambiguous. The instructions needed to be sufficiently clear such that if the rare threat event was observed, it was unambiguously classifiable as a target threat event. An earlier challenge to the interpretation of inattentional blindness was the possibility that participants may have observed the unexpected event (e.g. a gorilla or a surgical complication), but because that event lay outside of their provided task set, they were unsure what to do with that information and thus did not encode or report its presence (Memmert, 2010). The present task was designed to avoid that limitation. The task set was threats to soldiers, defined as people carrying weapons. A person with a knife and a person with a rifle both unambiguously fall within that category.

Yet the instructions also needed to be sufficiently ambiguous such that it was possible for an unexpected event to occur. For this reason, the instructions did not clarify that there were actually two target types: frequent threats (with a knife) and rare threats (with a gun). This ambiguity was strengthened by the frequency with which the knife subtype occurred in the two initial (training) trials. These trials exclusively involved threats holding knives, thereby implying that the class “armed” was actually operationalised as “wielding a knife” in the present game. In this manner the expectation of a person carrying a rifle was manipulated to be low, without directly stating that no other threat types would be present.

These instructions and the trial sequencing were expected to balance participants' need to know the task requirements against the well documented influence of expectancy on inattentional blindness (Mack & Rock, 1998). In brief, inattentional blindness is only observed when moment-to-moment expectancy is low. The strength of the expectancy effect on inattentional blindness was evident in the Experiment 1 results; after the first noticing of a rare threat character, participants were significantly more likely (approaching 100%) to notice the uncued rare threats on subsequent appearances. The fact that a sizeable proportion of participants noticed the rare threat in the uncued conditions in all three experiments and the Experiment 3 uninformative cue condition attests to the fact that expectancy was not excessively high on the critical condition of each experiment.

Given the large influence of expectancy on noticing rates in Experiment 1, it is tempting to conclude that this is the sole reason detection rates in the full-attention trials were so high and exceeded that in the critical trial. Expectancy was likely to be high during the full-attention trials for two reasons. The participant may have noticed the rare threat in their critical trial(s). Secondly, removing the "search for frequent threats" task and reiterating the "search for any threats" instruction strongly implies other possible threat types. The latter would make the full-attention more akin to a divided-attention trial (Mack & Rock, 1998; White et al., 2018).

However, it is important to moderate this conclusion given that two other changes happened simultaneously with the increased expectancy in the full-attention trials. First, the elimination of the frequent threat search reduced cognitive load. It therefore remains possible that the increase in detection rates during the full attention condition was brought about by freed cognitive resources facilitating detection of the rare threat. However, on balance this interpretation appears unlikely given the complete lack of evidence anywhere else in this experimental program that reduced cognitive load increases detection of the rare threat.

Second, in the full-attention trial, the requirement to respond to the frequent threats was removed, so the onscreen cues became task-irrelevant (for those in the informative cue conditions). The presence of valid cues did impact rare threat detection rates in Experiments 2 and 3 (see also uninformative cues in Experiment 2), so there is some reason to expect the change in status of the cues modified detection rates of the rare threat. However, this explanation offers no account of the comparable (and sharp) increase in detection rates in the full-attention trial, relative to the critical trial, in the no cue condition see in Experiments 2 and 3.

Therefore, expectancy offers the most comprehensive and parsimonious explanation for the sharp increase in detection rates between the critical and full-attention trials across experiments. It additionally explains the relatively high detection rates for the second and subsequent presentations of the rare threat in Experiment 1. But equally it appears likely that merely rendering the informative cues task irrelevant (in the informative cue conditions in Experiments 2 and 3) may have also facilitated detection of the rare threat in the full attention condition.

Summation

After reviewing the above hypotheses in light of the three experiments' findings, some broad trends emerged. Critically, no one hypothesis explains all of the observed data. Instead, the data are cumulatively best explained by a small set of hypotheses. Important questions remain concerning which factors are most important in shaping the impact of cues on inattentional blindness in different tasks or situations.

The hypothesis which received the least support was the cognitive load hypothesis, which aligns with findings reported in Mack and Rock (1998). The cognitive load hypothesis remains desirable due to its intuitive explanation of expected benefits from AR cueing, and its partial support from the literature (e.g., Hughes-Hallett et al., 2015; Simons & Chabris, 1999), despite this it provides the least likely explanation for any results obtained. Indeed, its core prediction that informative cueing would facilitate threat detection and thereby allow for greater sensitivity to unexpected threats was in the opposite direction to the observed data which supports findings showing a cost to cueing (e.g., Yeh & Wickens, 2001).

Expectancy clearly played a significant role in the rates of inattentional blindness observed. Experiment 1 presented participants with six rare threat events sequentially. While this effect increased experimental power for a given number of subjects (by using a within-subjects design and allowing for multiple measurements per person), it maximally exposes the design to any influence of expectancy. After participants see the rare event once, they become aware of its possible presence, and could be expected to adjust their search behaviour accordingly. This is precisely what was seen, with participants’ noticing rates for the rare event primarily being driven by whether they had observed the event before. With this large effect of expectancy in place, no effect of cueing was observed. A similar trend was seen for the full-attention conditions in all three experiments. Once participants had reason to expect that a search target existed beyond the frequent threats, detection rates of the rare threat approached ceiling. Consequently, the main data sets available to discriminate between the three hypotheses are performance on the critical trials, and primarily in Experiments 2 and 3, in which the rare threat was only shown once.

Of these conditions and trials, the attentional tunnelling approach generally provides a good account. It accurately describes the data of 5 of the 6 conditions in the experiments. The informative cue conditions in Experiments 2 and 3 both demonstrated facilitated responding to the frequent threats, and impaired detection of the rare threat, relative to the no cue condition, as predicted by the account. Further, the general ignoring of the uninformative cue in Experiment 3, such that responding matched that of the no cue condition, was well accounted for by the attentional tunnelling account. These outcomes align well with the AR literature, such as Wickens and Alexander (2009) which revealed attentional tunnelling via participants prioritising overt attention towards cued stimuli over uncued stimuli. The primary shortcoming of the attentional tunnelling account was its inability to explain the performance of the uninformative condition in Experiment 2. So how could this condition best be understood?

One approach is to consider the role of visual clutter. As predicted by the visual clutter account, rare threat detection seemed to be primarily impacted by whether or not a cue was present, rather than by its validity in predicting threats. Put another way, both cueing conditions saw reductions in detection rates for the rare event. Further, the presence of uninformative cues actively slowed detection of frequent threats in the uninformative condition in Experiment 2. Finally, global search behaviour as indexed by the breadth of the distribution of gaze points, also showed a restriction of search to the most probable target locations (the positions of the friendly soldiers). Each data point is indicative of a role for visual clutter, in which the presence of uninformative cues impaired deliberative, endogenous search.

A further question then arises as to why visual clutter played a crucial role in understanding the data pattern in the screen-based task format (Experiment 2) but not the VR-based format (Experiment 3). As discussed in Chapter 4, we speculate that the key differences relate to the expected improved visual realism of VR compared to screen based equivalents (Brookes et al., 2020; Kisker et al., 2019). More specifically within the present experiment the field of view provided by the VR display format.

The scene and characters (the Unity “world”) were held constant across experiments. The primary difference was in the manner in which that world was viewed. Although it is not immediately obvious, the traditional camera behaviour in video game formats – which was used in the screen-based Experiment 2 – has a subtle but notable “fish-eye” element. That is, more visual degrees of the world are viewable (typically approximately 100 degrees) than would ordinarily be possible in proportion of the participant’s real world visual field subtended by the screen (approximately 50 degrees for a 24” screen viewed at 60cm distance). In contrast, VR represents approximately the same field of view (approximately 100 degrees) but does so by wrapping that display around 100 visual degrees of their actual visual field. Consequently, the default display method of the same scene is less dense – or less cluttered – in the VR representation than the screen-based representation of the same scene. It is for this reason that we believe the VR format (Experiment 3) revealed less of an influence from visual clutter than observed in the screen-based format (Experiment 2).

A pressing question arising from this distinction is which format better generalizes to the field. More generally, this explanation raises questions as to the degree to which key insights regarding attention derived from screen-based lab tasks – even those using elaborate three-dimensional animations like those used here – can be safely expected to generalize to the experience of non-screen based, analogue visual perception. Consideration of the likely applied functionality may provide guidance. If real-world AR visual cues are provided via goggles or glasses then a VR system would likely provide a more valid approximation. Alternatively, screen-based may be more valid if real-world cues are provided through a weapons targeting system (e.g., rifle scope).

Implications, Limitations, and Future Research

The present results add to a growing body of literature showing that AR visual cues should be implemented in real-world settings with care and consideration. While the act of introducing cues will likely improve detection for a cued object, these benefits are likely to incur unforeseen costs. The implications of the present research were considered with respect to: (1) the potential for cueing to induce or amplify inattentional blindness (the thesis title); (2) the specific operational search context modelled in this task; and (3) virtual reality modelling of the task.

AR and Inattentional Blindness

The founding research question was whether, or to what degree, do visual cues induce inattentional blindness? This question was empirically answered: yes, cueing increases the risk of inattentional blindness. This answer holds at least in the military search task modelled, but we see no obvious reasons why this conclusion would not generalize to other search or detection tasks aided by visual overlays. The results add to a growing body of literature reporting counter-intuitive outcomes, whereby providing AR visual cues in operational search contexts may improve some aspects of search but reduces overall performance measures, particularly for uncued targets or other concurrent tasks (e.g., Dixon et al., 2013). When reliable cues are provided, operators are likely to improve performance on the AR-supported task but at the cost of poorer performance on other unsupported, simultaneous tasks.

Beyond the general bias towards valid, salient cues there was also some evidence (particularly the uninformative condition in Experiment 2) that the mere presentation of unreliable cues may reduce overall search performance due to increasing visual clutter. This is important because in complex and dynamic operational search environments, the reliability of a cueing system is likely to vary significantly. Literature suggests a 60% reliability limit for cognitively demanding contexts, whereby people will ignore a cue with less than 60% reliability (e.g., Wickens et al., 2006). The reliability levels of the 20 studies synthesised by Wickens and Dixon (2007) show no studies examined reliability levels below 50%. Yet the present data (Experiment 2) suggest that the presence of non-contingent cues can also impact search, even when the cues were known to be unreliable and people sought to ignore them, rather than use them.

Crucially, the dependent variable in many decision aid studies is people's reported reliance on decision aids or the higher order judgments generated in the presence of an aid. By contrast, the present task examined facilitation of an elemental search task that could be completed without an aid. The question here was not whether complex judgments or classifications were supported, but whether a series of time-sensitive elementary actions were facilitated or impaired by the presence of cues. In this context, unreliable cues were detrimental to performance, at least in the screen-based format. The generality of this finding to other actions and judgments, and other degrees or reliability, remains a question for further research.

It is interesting that participants could not simply ignore the uninformative cues in Experiment 2. Participants were told that the cues were uninformative in advance, the cues were reliably uninformative throughout, and by the time of the critical trial, participants had had minutes of exposure to these uninformative cues in which they could have practiced ignoring them. It is perhaps notable that the cues were uninformative rather than invalid. That is, approximately 33% of the characters were frequent threats. To instantiate a zero contingency between cueing and threat status, that meant that 33% of the cued characters were actually threats. That is, the base rate of being a threat was held constant for cued and uncued characters. This design choice was chosen because if cues were to be made perfectly invalid, such that only non-threat characters were cued, the cue would actually constitute a meaningful, predictive signal. That is, an optimal strategy would be to notice the presence of the cue, and then direct search elsewhere rather than to simply ignore the visual cues.

While the uninformative cues in Experiment 2 were objectively, statistically uninformative, it remains possible that people were particularly sensitive to the occasions in which a cued character performed a threat action and less sensitive (or perhaps were less likely to detect) the instances in which uncued characters performed threat actions. In this sense, the cues may have been perceived as partially predictive, thereby maintaining some degree of endogenous attentional prioritization, despite their objective irrelevance. Such an explanation might also explain some of the difference in the performance of the uninformative group between Experiments 2 and 3. In the VR condition, a smaller percentage of the scene would be available for high precision inspection in or near the foveal region than in the screen-based format. Consequently, participants in the uninformative condition may have been less likely to accidentally observe cued characters performing a threat action, thereby leading to less instances of reinforcement for attending to cued characters, and in turn making it easier to ignore the objectively uninformative cues. The present design limits the ability to test such hypotheses, but extending Experiments 2 and 3 with an invalid cue control condition would help. Inclusion of such a condition would overcome a limitation of the present design, which is unable to reveal if the unexpected attention towards uninformative cues was due to those cues retaining some small value, or due to other design features of the experiment. This too remains for future research.

Primary Search Task

As documented throughout many sections of this thesis the primary search task is a critical element in inattentional blindness paradigms, but also the primary task design for specific experiments garners much debate and investigation within the literature (e.g., Cartwright-Finch & Lavie, 2007; Mack & Rock, 1998; Simons & Jensen, 2009). For the present research a priority was placed on ensuring the instructions given to participants would be adequate to guide them in performing the primary task but those same instructions should not deceive or guide them away from noticing the unexpected object. Hence the instructions were to search for any threat and threats were defined as weapons. The primary search task involved searching for knives, whilst the unexpected object was a character carrying a rifle. This contrasts to recurring criticisms of experiments such as the gorilla experiment (Simons & Chabris, 1999) where the primary task is counting basketball passes but participants are judged on their noticing or not an unexpected object unrelated to counting basketball passes, i.e., individual in a gorilla suit (c.f., Hughes-Hallett et al., 2015; Yeh & Wickens, 2001).

The other design decision related to the primary search task was the rate of frequent threat appearances. Too low a rate and the task would be too easy and inattentional blindness was unlikely, too high a rate and the task would simply be too difficult. The intent of suitably adjusting the task difficulty (via frequent threat appearance rate) was not to investigate if inattentional blindness would occur but rather to ensure it would occur to enable investigating if cueing modulated inattentional blindness.

Predominantly, the primary search task achieved its design goals. Inattentional blindness was generated, and the modulation of it by cueing was able to be investigated. However, there were some limitations. For example, the Experiment 3 informative cue condition only had one noticer. Such low noticing rates made statistical comparisons between conditions inappropriate.

An additional limitation was that despite the gaze metrics in the present research proving valuable, the primary task design involved monitoring four centrally located friendly characters as well as the targets and distractors. Thus, a frequent threat event was the product of the conjunction of two characters, a friendly soldier and the attacker, moving in close proximity to each other. This jointly defined search target complicated the inferences that could be drawn from gaze behaviour: was a participant gazing at a friendly soldier, or anticipating the future trajectory of a threat character, or both? This ambiguity was particularly challenging because a side effect of the primary task of clicking on the frequent threats was that when they attacked the soldiers it led to clusters of characters and fixations centred on the soldiers' positions. This lumpiness in the distribution of character location and gaze point limited the experimental power to determine if fixations were a product of participants inspecting and classifying characters, monitoring the position of frequent threat characters, or gazing at a potential threat in service of performing the clicking action on that character.

To better investigate how much attention is given to uninformatively cued characters, it would be desirable to minimise the potential for clusters of fixations and joint or overlapping regions of interest. Any such change should make it clearer if a character is being looked at for targeting purposes, is being accidentally fixated upon due to its proximity to the friendly soldiers or is being attended to as part of the search strategy.

One suggestion is to remove the friendly soldiers from the scene altogether, and instead participants would be asked to identify any character that revealed a weapon irrespective of their location. These threat events could then be dispersed widely around the scene, facilitating the interpretation of gaze data. The use of threats moving towards and intersecting with a friendly soldier was intended as an analogue of the edge collisions people counted in the widely known "bouncing shapes" inattentional blindness task (Most et al., 2001). But there is no compelling reason that other, non-collision events could not be used to define a search task (abrupt onsets were used to excellent effect in Mack & Rock's original studies). This suggestion would also resolve a potential criticism of the current experiments whereby it could be suggested that participants searched for the knife attack motion rather than for the knife object or proximity to a friendly soldier. However, any change would need to be carefully considered with respect to its potential to mislead or deceive, both for ethical reasons (ethics research approval panels frown upon deception) and for the conceptual confusion that can arise when distinguishing inattentional blindness from certain kinds of intentional misdirection (e.g., as used in magic tricks) or change blindness (Jensen et al., 2011; Memmert, 2010; Yao et al., 2019).

Dismounted Soldier

Certain aspects of the video game experiment create a reasonable approximation to a potential real-world overwatch scenario (Monroe, 2014). However, other aspects were necessarily more contrived. These first experiments confirm that AR-induced inattentional blindness is possible within these contexts but do not address how well the findings might generalise to the real world of AR-supported soldiers performing overwatch.

One initial consideration is that the present level of technology and the variability of contexts that soldiers operate in make it unlikely that AR systems will reliably identify threats on the battlefield in the near term. Real-world reliability rates are difficult to obtain, particularly from military sources. In simpler, static image classification tasks, machine vision can perform spectacularly well. For example, a tyre manufacturer's case study report shows 99.9% reliability in identifying manufacturing defects (intel, 2023). This rate is obtained in the perfect environment of a factory with clean images taken from multiple fixed imaging systems, processed by an industrial PC trained to search for specific defects. It would appear unlikely that a jostled, lightweight head camera mounted on a soldier, subject to environmental conditions and relying on local processing resources could achieve the same level of precision when classifying a suspected adversary from hundreds of metres away.

Instead, it appears more likely that systems will be able to visually represent the location of known signals or features (such as using the location of nearby Identifying Friend or Foe, IFF, signals to localise friendly soldiers). Any visual representation of this kind would define a target that likely does *not* require further investigation. Research has indicated that such systems aid performance (Bryant & Smith, 2013), however, whether such no-go visual cues impact on inattentional blindness remains an open question.

This discrepancy between highlighting enemies versus the more likely scenario of highlighting friends represents an important limitation in generalisability of the present experiments. Future research should consider a reversal of the highlighting to reflect the most likely current and near-future scenario of IFF systems highlighting friend rather than foe.

Finally, the present research used a university student participant pool, raising a concern of generalisation. None of our participants had performed real-world overwatch duties, however, the university student pool demonstrated a suitable level of skill and engagement in the experiments. Furthermore, the practice trials provided participants time to familiarise themselves with the task, but they were untrained in search techniques and did not report using particular strategies or techniques to remain vigilant. Would well-trained individuals realise the same levels of inattentional blindness?

Research such as Drew et al. (2013) would suggest that even experts are not immune to inattentional blindness, as they reported 83% of expert radiologists experienced inattentional blindness in their study. Although the present project intended to test the present task using military personnel, funding limitations and military deployment to natural disasters resulted in the armed forces withdrawing access before testing could be completed. So ultimately this question remained unanswered. Future work could resolve this limitation by training some participants in specific search techniques as a feasible substitute to accessing military operators. Alternatively, accessing other populations that have received search training may be viable, such as military cadets or Army reservists.

Virtual Reality

The informative cue condition results from Experiment 3 (VR) mirrored those of Experiment 2. However, inattentional blindness rates were reduced in Experiment 3 for the uninformative cue condition. This reduction was interpreted as a result of perceptual differences between the screen and VR experiments. However, there were no specific design elements in this series of experiments that directly tested our interpretation. Furthermore, the percentage of the visual field covered by display was not the only difference between Experiment 2 and 3 displays formats. The most obvious difference was the use of black, opaque “curtains” (see Chapter 4) which narrowed the VR scene to approximately the width of the scene shown in the screen-based task. So, despite participants, in theory, having an increased ability to move their heads to investigate the visual scene, the curtains removed the capacity to do so. Without these curtains, it was possible for a participant to simply point their head in the wrong direction and thereby fail to detect any threats in the scene. Such behaviour need not be controlled for in the screen-based task, as the camera position was fixed. Yet it also remains possible that the inclusion of this control device may also have impacted upon both the effects of cueing and the prevalence of inattentional blindness.

The comparison of Experiment 2 and 3 results would suggest it is unlikely that these game design changes for VR had a significant impact as the uncued, informatively cued and full-attention trials all showed similar patterns of results to Experiment 2. However, as one of the core conclusions from the present experiments was premised on the assumption that VR reduced visual clutter, it would be advantageous to confirm this through further research. Such alternative accounts cannot empirically be dismissed representing a limitation to interpreting the Experiment 3 results. Replicating the entire task without curtains would aid in removing this limitation however, careful consideration would need to be given to experiment and scene design. Both the screen-based and VR experiments would need to be modified to enable suitable comparisons between the two, whilst also resolving the underlying issues that necessitated the curtains for the present experiment.

SSVEP/EEG

Earlier in this section a limitation was noted about the primary search task and its impact on eye gaze measure. That same limitation also relates to the SSVEP measures in Experiment 2 and 3. In summary, the need for participants to focus on the four friendly soldiers as part of the primary search task resulted in overt attention towards the cued characters for targeting. Leading to the overt attention recorded within the SSVEP signal likely overpowering the covert signal (Walter et al., 2012). Experiment designs that more clearly segregate the characters on the screen will not only aid the eye gaze measures but also decrease the chance of overt signals overpowering covert signals in the SSVEP measure.

In addition, during the VR experiment it was suspected that the weight and movement of the VR headset potentially increased the number and magnitude of movement artefacts appearing in the SSVEP signal. The evidence suggested that the VR and EEG headsets did not physically interfere with each other (See Chapter 4), however it remains unclear if the SSVEP signal was tainted by the VR headset movements shifting the EEG cap during the experiment. Such movements would likely have been physically imperceptible, but significant relative to EEG data collection best practices. Whilst SSVEP is considered robust to movement artefacts, and techniques to remove these artefacts exist (Luck & Kappenman, 2017; Norcia et al., 2015), the literature is generally referring to muscle movement artefacts and not the electrode/cap movement artefacts suspected in the present experiment. It is therefore suggested that future experiments using both VR and EEG should investigate improved fitment or mounting options to eradicate potential physical interference between the two systems.

Multi-methods Approach

The present research included measurements involving behavioural, eye-tacking, and EEG measures. This combination was expected to allow a triangulation of results, and it was also hoped the inclusion of EEG would add to the literature in areas such as the influence of cueing on covert attention. However, exploiting such different methods meant compromises were inevitably necessary.

As discussed in detail in the preceding section, despite best efforts and initial positive experimenter feedback it is suspected that the VR headset prevented implementing EEG/SSVEP best practices and the EEG signal likely suffered as a result. The Biosemi EEG system employed for the study uses active electrodes which tolerate high impedances (BioSemi B.V., n.d.), helping compensate for poorer scalp connections. However, it is reasonable to assume that such a system was not designed with simultaneous fitment of VR headsets in mind. Future research with the potential for physical interference between devices or experiments requiring greater than usual head movements from participants should investigate techniques such as ambulatory-EEG (Kwak et al., 2017). Such techniques are designed to cope with less than perfect EEG data collection scenarios, and may provide insights into better options or analyses that can be adapted to the present research design.

In addition to the likely VR and EEG headset conflicts, the cue design was another example of EEG requirements influencing design elements. The cues were designed to flash at a rate of 15Hz, ensuring the rate was within the 3-20Hz frequency range commonly used for SSVEP (Norcia et al., 2015) due to optimal SSVEP signal-to-noise ratios being more likely (Ladouce et al., 2022). The dilemma is that a 15Hz flashing rate is clearly visible to the human eye and likely to increase cue salience. So, a higher rate was excluded due to the risk of poor signal-to-noise ratios and a lower rate was excluded due to the increased potential for participant discomfort (Ladouce et al., 2022). However, as the cues were designed to be highly salient it was decided that the 15Hz rate was an acceptable compromise.

Despite these compromises, and the others noted throughout the thesis, the three methods selected (behavioural, eye tracking, and SSVEP) worked well together. Importantly, most behavioural responses were unsurprising when compared to other AR visual search literature, suggesting success in that domain. Furthermore, eye gaze measures mostly aligned with the behavioural responses suggesting those two domains supported each other. Finally, despite SSVEP responses appearing underpowered (see SSVEP/EEG section above) they trended towards aligning with the eye-gaze data. This final outcome suggesting potential for this unique application of SSVEP if further development is conducted.

Whilst it is beyond the scope of this section to repeat and revisit all the individual design choices documented throughout this thesis, it is important for the reader to be cognisant that the multi-methods approach chosen did involve compromise. Any future research should not only consider the specific limitations and ideas above but also consider the influence of design decisions on each of the multi-methods selected. For example, it may be inappropriate in some studies to retain SSVEP unless resolving SSVEP specific issues can be prioritised.

Conclusion

This thesis investigated whether AR visual cues would help or hinder operational search in a dismounted soldier performing overwatch. AR visual cues improved both hit rates and response times for reliably cued targets, but at the cost of poorer search outcomes for rare uncued targets. There was evidence of attentional tunnelling towards the task-relevant cues, preventing operators from adequately inspecting all objects in the scene. Further, Experiment 2 provided evidence of uninformative cues impairing threat detection for both frequent and rare threats, due to the visual clutter imposed by large, salient but ultimately unhelpful cues. The present experimental series also found subtle, but likely important differences in the impact of cueing on inattentional blindness in screen-based versus VR-based displays. This observation raises questions about the generality of screen-based lab findings for real-world applications, and indeed whether the present VR-based tasks would generalize to more realistic animations or real-world settings. Cumulatively the experiments showed that cueing can exert both costs and benefits to performance. Therefore, development of safe, productive AR systems requires consideration of both the costs to search sensitivity as well as the facilitatory effects of onscreen cues.

References

Adobe. (2020). *Adobe Fuse*. https://www.adobe.com/wam/fuse.html

Agresti, A. (2019). *An introduction to categorical data analysis* (3rd. ed.). John Wiley & Sons, Inc.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proc Natl Acad Sci USA*, *108*(25), 10367-10371. https://doi.org/10.1073/pnas.1104047108

Andri et mult. al., S. (2022). DescTools: Tools for Descriptive Statistics. In.

Beanland, V., & Pammer, K. (2010). Looking without seeing or seeing without looking? Eye movements in sustained inattentional blindness. *Vision Res*, *50*(10), 977-988. https://doi.org/10.1016/j.visres.2010.02.024

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). (Ben-Shachar et al.) effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*, 2815. https://doi.org/10.21105/joss.02815

BioSemi B.V. (n.d.). *Biosemi*. www.biosemi.nl

Boneau, C. A. (1960). The effects of violations of assumptions underlying the test. *Psychol Bull*, *57*, 49-64. https://doi.org/10.1037/h0041412

Boyd, J. R., & Hammond, G. T. (2018). A Discourse on Winning and Losing.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*(2), 455-463. https://doi.org/10.3758/s13428-019-01242-0

Bryant, D. J., & Smith, D. G. (2013). Impact of blue force tracking on combat identification judgments. *Hum Factors*, *55*(1), 75-89. https://doi.org/10.1177/0018720812451595

Calibration - Tobii Pro SDK documentation. (2021). https://developer.tobiipro.com/commonconcepts/calibration.html

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition*, *102*(3), 321-340. https://doi.org/10.1016/j.cognition.2006.01.002

Chabris, C. F., Weinberger, A., Fontaine, M., & Simons, D. J. (2011). You Do Not Talk about Fight Club if You Do Not Notice Fight Club: Inattentional Blindness for a Simulated Real-World Assault. *i-Perception*, *2*(2), 150-153. https://doi.org/10.1068/i0436

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Routledge. https://doi.org/https://doi.org/10.4324/9780203771587

Cohen, M. X., & Gulbinaite, R. (2017). Rhythmic entrainment source separation: Optimizing analyses of neural responses to rhythmic sensory stimulation. *Neuroimage*, *147*, 43-56. https://doi.org/10.1016/j.neuroimage.2016.11.036

Cooke, G. (2019). *The Future of Artificial Intelligence in Weapon Systems*. https://www.defensemedianetwork.com/stories/the-future-of-artificial-intelligence-in-weapon-systems/

Cumming, G. (2014). The new statistics: why and how. *Psychol Sci*, *25*(1), 7-29. https://doi.org/10.1177/0956797613504966

Dell Inc. (2020). *Laptops, Monitors, Computers & Storage Solutions | Dell Australia*. https://www.dell.com/en-au

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Devagiri, J. S., Paheding, S., Niyaz, Q., Yang, X., & Smith, S. (2022). Augmented Reality and Artificial Intelligence in industry: Trends, tools, and future challenges. *Expert Systems with Applications*, *207*. https://doi.org/10.1016/j.eswa.2022.118002

Dixon, B. J., Daly, M. J., Chan, H., Vescan, A. D., Witterick, I. J., & Irish, J. C. (2013). Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical Endoscopy*, *27*(2), 454-461. https://doi.org/10.1007/s00464-012-2457-3

Dixon, S. R., & Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(3), 474-486. https://doi.org/10.1518/001872006778606822

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: are automation false alarms worse than misses? *Hum Factors*, *49*(4), 564-572. https://doi.org/10.1518/001872007X215656

Drew, T., Vo, M. L., & Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattentional blindness in expert observers. *Psychol Sci*, *24*(9), 1848-1853. https://doi.org/10.1177/0956797613479386

Duarte, M., & Zatsiorsky, V. M. (2002). Effects of body lean and visual information on the equilibrium maintenance during stance. *Experimental Brain Research*, *146*(1), 60-69. https://doi.org/10.1007/s00221-002-1154-1

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice* (3rd ed.). Springer.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol Rev*, *96*(3), 433-458. https://doi.org/10.1037/0033-295x.96.3.433

Eisma, Y. B., Borst, C., Paassen, R. V., & Winter, J. D. (2020). Augmented Visual Feedback: Cure or Distraction? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082092460. https://doi.org/10.1177/0018720820924602

Ekelund, M., Fernsund, H., Karlsson, S., & Mac Giolla, E. (2022). Does Expertise Reduce Rates of Inattentional Blindness? A Meta-Analysis. *Perception*, *51*(2), 131-147. https://doi.org/10.1177/03010066211072466

Endsley, M. R., & Jones, W. M. (1997). Siatuation awareness information dominance & Information warfare.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fish, L., & Scharre, P. (2018). *The Soldier's Heavy Load* (Super Soldiers, Issue. https://www.cnas.org/publications/reports/the-soldiers-heavy-load-1

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030-1044. https://doi.org/10.1037/0096-1523.18.4.1030

Frame, M. E., Warren, R., & Maresca, A. M. (2019). Scanpath comparisons for complex visual search in a naturalistic environment. *Behavior Research Methods*, *51*(3), 1454-1470. https://doi.org/10.3758/s13428-018-1154-0

Funder, D. C., & Ozer, D. J. (2020). Corrigendum: Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *3*(4), 509-509. https://doi.org/10.1177/2515245920979282

Galster, S. M., Bolia, R. S., & Parasuraman, R. (2016). Effects of Information Automation and Decision-Aiding Cueing on Action Implementation in a Visual Search Task.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, *46*(3), 438-442. https://doi.org/10.1177/154193120204600347

Geospatial World. (2014). *Bringing Augmented Reality Systems on the Battlefield*. Geospatial World. https://www.geospatialworld.net/article/bringing-augmented-reality-systems-on-the-battlefield/

Gilchriest, J. A. (2011). A method for quantifying visual search scanpth efficiency. *Functional Neurology, Rehabilitation, and Ergonomics*, *1(2)*, 181-196. https://www.proquest.com/docview/1726342587/fulltext/2B21F83194434677PQ

Gilchrist, I. (2012). Saccades. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements*. https://doi.org/10.1093/oxfordhb/9780199539789.001.0001

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychol Sci*, *18*(1), 88-94. https://doi.org/10.1111/j.1467-9280.2007.01853.x

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. R. E. Krieger Pub. Co.

Griffiths, O., Gwinn, O. S., Russo, S., Baetu, I., & Nicholls, M. E. R. (2021). Reinforcement history shapes primary visual cortical responses: An SSVEP study. *Biological Psychology*, *158*, 108004. https://doi.org/10.1016/j.biopsycho.2020.108004

Hancock, P. A., Mercado, J. E., Merlo, J., & Van Erp, J. B. (2013). Improving target detection in visual search through the augmenting multi-sensory cues. *Ergonomics*, *56*(5), 729-738. https://doi.org/10.1080/00140139.2013.771219

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: evidence from search efficiency and eye movements. *J Vis*, *9*(1), 32 31-38. https://doi.org/10.1167/9.1.32

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(6693), 575-577. https://doi.org/Doi 10.1038/29068

Hughes-Hallett, A., Mayer, E. K., Marcus, H. J., Pratt, P., Mason, S., Darzi, A. W., & Vale, J. A. (2015). Inattention blindness in surgery. *Surg Endosc*, *29*(11), 3184-3189. https://doi.org/10.1007/s00464-014-4051-3

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: effects with single- and combined-feature targets. *J Exp Psychol Gen*, *118*(3), 258-279. https://doi.org/10.1037//0096-3445.118.3.258

Hutchinson, B. T., Pammer, K., Bandara, K., & Jack, B. N. (2022). A tale of two theories: A meta-analysis of the attention set and load theories of inattentional blindness. *Psychological bulletin*, *148*(5-6), 370-396. https://doi.org/10.1037/bul0000371

Hutchinson, B. T., Wilkinson, N., Robertson, G., Budd, A., Nicholls, M. E. R., & Griffiths, O. (2023). An investigation of inattentional blindness using gaze and frequency tagging. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xhp0001143

IBM. (2020). *Maximo Application Suite - Overview*. https://www.ibm.com/products/maximo

intel. (2023). *Computer vision solution for industrial PCs helps improve defect detection*. https://www.intel.com/content/dam/www/public/us/en/documents/case-studies/iotg-ipc-tire-case-study-fv.pdf

JASP Team. (2023). *JASP*. In (Version 0.17.2) https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability* (3rd. ed.). Clarendon Press, Oxford.

Jensen, M. S., Yao, R., Street, W. N., & Simons, D. J. (2011). Change blindness and inattentional blindness. *Wiley Interdiscip Rev Cogn Sci*, *2*(5), 529-546. https://doi.org/10.1002/wcs.130

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1-19. https://doi.org/10.1007/s10648-010-9150-7

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. In.

Kisker, J., Gruber, T., & Schone, B. (2019). Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychol Res*. https://doi.org/10.1007/s00426-019-01244-9

Koivisto, M., Hyona, J., & Revonsuo, A. (2004). The effects of eye movements, spatial attention, and stimulus features on inattentional blindness. *Vision Res*, *44*(27), 3211-3221. https://doi.org/10.1016/j.visres.2004.07.026

Kothe, C. A. E., & Jung, T. P. (2016). *Artifact removal techniques with signal reconstruction* (U.S. Patent No. https://patents.google.com/patent/US20160113587A1/en

Kuhn, G., Tatler, B. W., Findlay, J. M., & Cole, G. G. (2008). Misdirection in magic: Implications for the relationship between eye gaze and attention. *Visual Cognition*, *16*(2-3), 391-405. https://doi.org/10.1080/13506280701479750

Kwak, N. S., Muller, K. R., & Lee, S. W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLoS One*, *12*(2), e0172578. https://doi.org/10.1371/journal.pone.0172578

Ladouce, S., Darmet, L., Torre Tresols, J. J., Velut, S., Ferraro, G., & Dehais, F. (2022). Improving user experience of SSVEP BCI through low amplitude depth and high frequency stimuli design. *Sci Rep*, *12*(1), 8865. https://doi.org/10.1038/s41598-022-12733-0

Le Pelley, M., Beesley, T., & Griffiths, O. (2016). Associative Learning and Derived Attention in Humans. In *The Wiley handbook on the cognitive neuroscience of learning* (pp. 114-135). Wiley Blackwell. https://doi.org/10.1002/9781118650813.ch6

Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When goals conflict with values: counterproductive attentional and oculomotor capture by reward-related stimuli. *J Exp Psychol Gen*, *144*(1), 158-171. https://doi.org/10.1037/xge0000037

Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res*, *48*(5), 635-654. https://doi.org/10.1016/j.visres.2007.12.009

Lo, S.-Y., & Lai, C.-Y. (2023). Investigating how immersive virtual reality and active navigation mediate the experience of virtual concerts. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-35369-0

Luck, S. J., & Kappenman, E. S. (2017). Electroencephalography and event-related brain potentials. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 74–100). Cambridge University Press.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics*, *44*(3), 229-243. https://doi.org/10.1080/00140130010000901

Mack, A., & Rock, I. (1998). *Inattentional blindness*. The MIT press.

MATLAB. In. (Version R2019b) The MathWorks Inc. https://au.mathworks.com/

Memmert, D. (2010). The gap between inattentional blindness and attentional misdirection. *Conscious Cogn*, *19*(4), 1097-1101. https://doi.org/10.1016/j.concog.2010.01.001

Merritt, S. M. (2011). Affective processes in human-automation interactions. *Hum Factors*, *53*(4), 356-370. https://doi.org/10.1177/0018720811411912

Monroe, S. R. (2014). *Overwatch*. https://www.dvidshub.net/image/1302440/overwatch

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: sustained inattentional blindness and the capture of awareness. *Psychol Rev*, *112*(1), 217-242. https://doi.org/10.1037/0033-295X.112.1.217

Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological science*, *12*(1), 9-17.

Mullen, T. (2012). CleanLine EEGLAB plugin. *San Diego, CA: Neuroimaging Informatics Toolsand Resources Clearinghouse (NITRC)*.

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*(1), 1-19. https://doi.org/10.1017/s0140525x12003214

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *J Vis*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230-253. https://doi.org/10.1518/001872097778543886

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. *Data Brief*, *25*, 104101. https://doi.org/10.1016/j.dib.2019.104101

Pitts, M. A., Martinez, A., & Hillyard, S. A. (2012). Visual processing of contour patterns under conditions of inattentional blindness. *J Cogn Neurosci*, *24*(2), 287-303. https://doi.org/10.1162/jocn_a_00111

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, *3*(3), 179-197.

R Development Core Team. (2020). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychol Rev*, *74*(1), 71-80. https://doi.org/10.1037/h0024109

Richards, A., Hannon, E. M., & Vitkovitch, M. (2012). Distracted by distractors: eye movements in a dynamic inattentional blindness task. *Conscious Cogn*, *21*(1), 170-176. https://doi.org/10.1016/j.concog.2011.09.013

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*.

Schöne, B., Sylvester, R. S., Radtke, E. L., & Gruber, T. (2020). Sustained inattentional blindness in virtual reality and under conventional laboratory conditions. *Virtual Reality*. https://doi.org/10.1007/s10055-020-00450-w

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074. https://doi.org/10.1068/p281059

Simons, D. J., & Jensen, M. S. (2009). The effects of individual differences and task difficulty on inattentional blindness. *Psychonomic bulletin & review*, *16*(2), 398-403. https://doi.org/10.3758/pbr.16.2.398

Skinner, D. (2019). *The integrated digital soldier concept*. https://cove.army.gov.au/sites/default/files/the_integrated_digital_soldier_concept_v1.pdf

Speed, A. (2015). Visual Search in Operational Environments. In *Foundations of Augmented Cognition* (pp. 528-536). https://doi.org/10.1007/978-3-319-20816-9_51

Takeuchi, T. (1997). Visual search of expansion and contraction. *Vision Research*, *37*(15), 2083-2090. https://doi.org/10.1016/s0042-6989(96)00225-8

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychol (Amst)*, *135*(2), 77-99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2013). Feature-based attention: it is all bottom-up priming. *Philos Trans R Soc Lond B Biol Sci*, *368*(1628), 20130055. https://doi.org/10.1098/rstb.2013.0055

Tobii AB. (2015). Tobii Pro VR Integration based on HTC Vive HMD.

Tobii AB. (2020). *Improving your research with eye tracking since 2001 - Tobii Pro*. https://www.tobiipro.com

Treisman, A. (1985). Preattentive Processing in Vision. *Computer Vision Graphics and Image Processing*, *31*(2), 156-177. https://doi.org/Doi 10.1016/S0734-189x(85)80004-9

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn Psychol*, *12*(1), 97-136. https://doi.org/10.1016/0010-0285(80)90005-5

Unity. In. (2018). Unity Technologies. https://unity3d.com/

Unity. (2020). https://unity.com

van den Berg, R., Cornelissen, F. W., & Roerdink, J. B. (2009). A crowding model of visual clutter. *J Vis*, *9*(4), 24 21-11. https://doi.org/10.1167/9.4.24

VIVE VR System. (2022). https://www.vive.com/eu/product/vive/#vive-spec

Walter, S., Quigley, C., Andersen, S. K., & Mueller, M. M. (2012). Effects of overt and covert attention on the steady-state visual evoked potential. *Neurosci Lett*, *519*(1), 37-41. https://doi.org/10.1016/j.neulet.2012.05.011

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and Reliance on an Automated Combat Identification System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(3), 281-291. https://doi.org/10.1177/0018720809338842

Watson, P., Pearson, D., Theeuwes, J., Most, S. B., & Le Pelley, M. E. (2020). Delayed disengagement of attention from distractors signalling reward. *Cognition*, *195*, 104125. https://doi.org/10.1016/j.cognition.2019.104125

Weber, S., Schubert, R. S., Vogt, S., Velichkovsky, B. M., & Pannasch, S. (2018). Gaze3DFix: Detecting 3D fixations with an ellipsoidal bounding volume. *Behavior Research Methods*, *50*(5), 2004-2015. https://doi.org/10.3758/s13428-017-0969-4

White, R. C., Davies, M., & Aimola Davies, A. M. (2018). Inattentional blindness on the full-attention trial: Are we throwing out the baby with the bathwater? *Consciousness and Cognition*, *59*, 64-77. https://doi.org/10.1016/j.concog.2017.10.002

Wickens, C. D., & Alexander, A. L. (2009). Attentional Tunneling and Task Management in Synthetic Vision Displays. *The International Journal of Aviation Psychology*, *19*(2), 182-199. https://doi.org/10.1080/10508410902766549

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., Dixon, S. R., & Johnson, N. (2006). Imperfect Diagnostic Automation: An Experimental Examination of Priorities and Threshold Setting. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 210-214. https://doi.org/10.1177/154193120605000301

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychon Bull Rev*, *1*(2), 202-238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: trained observers miss rare targets, too. *J Vis*, *13*(3), 33. https://doi.org/10.1167/13.3.33

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Front Hum Neurosci*, *9*, 131. https://doi.org/10.3389/fnhum.2015.00131

Woodward, J., & Ruiz, J. (2023). Analytic Review of Using Augmented Reality for Situational Awareness. *IEEE Transactions on Visualization and Computer Graphics*, *29*(4), 2166-2183. https://doi.org/10.1109/tvcg.2022.3141585

Yao, R., Wood, K., & Simons, D. J. (2019). As if by Magic: An Abrupt Change in Motion Direction Induces Change Blindness. *Psychol Sci*, *30*(3), 436-443. https://doi.org/10.1177/0956797618822969

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: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 390-407. https://doi.org/10.1518/hfes.45.3.390.27249

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