

Naval Submarine Medical Research Laboratory

NSMRL/F2001/TR--2022-1404

October 31, 2022



Elucidating Visual Search Behavior and Resulting Performance in Low Prevalence Dynamic Search

by

Krystina Diaz, BA[†]
Margaret Wise, BA[†]
Sylvia Guillory, PhD
Jeffrey Bolkhovsky, PhD
Chad Peltier, PhD

[†]Krystina Diaz and Margaret Wise contributed equally to this work.

Approved and Released by:
M. H. JAMERSON, CAPT, MSC, USN
Commanding Officer
NAVSUBMEDRSCHLAB

Approved for public release: distribution unlimited.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> <i>OMB No. 0704-0188</i>							
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.											
1. REPORT DATE (DD-MM-YYYY) 31-10-2022		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) 7/2020 - 10/2022							
4. TITLE AND SUBTITLE Elucidating Visual Search Behavior and Resulting Performance in Low Prevalence Dynamic Search				5a. CONTRACT NUMBER							
				5b. GRANT NUMBER N0001422WX01547							
				5c. PROGRAM ELEMENT NUMBER 0602236N							
6. AUTHOR(S) Krystina Diaz, Margaret Wise, Sylvia Guillory, Jeffrey Bolkhovsky, Chad Peltier				5d. PROJECT NUMBER							
				5e. TASK NUMBER							
				5f. WORK UNIT NUMBER F2001							
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Submarine Medical Research Laboratory Box 900 Groton, CT 06349-5900				8. PERFORMING ORGANIZATION REPORT NUMBER NSMRL/F2001/TR--2022-1404							
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research One Liberty Center 875 N. Randolph Street, Suite 1425 Arlington, VA 22203-1995				10. SPONSOR/MONITOR'S ACRONYM(S) ONR							
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)							
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release: distribution unlimited.											
13. SUPPLEMENTARY NOTES											
14. ABSTRACT Visual search is an essential element of various professions, such as airport security screening, radiology, and sonar monitoring, in which accurate and efficient task performance is vital. Visual search performance is dependent on specific search conditions, such as how many targets are present, if targets are moving or stationary, and how long the targets are visible. Despite existing research on how conditions such as these individually influence search performance, the extent to which they interact—particularly target prevalence in different types of search environments—is unclear. Therefore, this study investigated performance and perceived cognitive demands of three common aspects of visual search: target prevalence (50% vs 10%), display type (static still-image vs. dynamic continuous scroll), and display speed (3.5 vs. 7 seconds). Participants (n = 556) recruited through Amazon Mechanical Turk were randomly assigned to one of eight experimental conditions created from the combinations of these search conditions. Participants were tasked with detecting known targets, “T”s, among a display of distractors, offset “L”s, and subsequently rating their perceived condition-specific cognitive task workload via the NASA Task Load Index (NASA-TLX). Though no significant interactions between search conditions were found, condition-specific effects were present. Lower target prevalence produced decreased hit rates, replicating the Low Prevalence Effect (LPE) (Wolfe et al., 2007), across static and dynamic displays. Additionally, NASA-TLX ratings implies observer blindness to actual performance decrements in low target prevalence searches. The persistence of the LPE highlights the need for LPE-mitigation methods in both types of search environments.											
15. SUBJECT TERMS visual search, low prevalence effect, dynamic search, simulated targets, search strategy, cognitive workload											
16. SECURITY CLASSIFICATION OF: <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr> <td style="width: 33%; padding: 2px;">a. REPORT</td> <td style="width: 33%; padding: 2px;">b. ABSTRACT</td> <td style="width: 33%; padding: 2px;">c. THIS PAGE</td> </tr> <tr> <td style="text-align: center; padding: 2px;">U</td> <td style="text-align: center; padding: 2px;">U</td> <td style="text-align: center; padding: 2px;">U</td> </tr> </table>			a. REPORT	b. ABSTRACT	c. THIS PAGE	U	U	U	17. LIMITATION OF ABSTRACT SAR		18. NUMBER OF PAGES 36
a. REPORT	b. ABSTRACT	c. THIS PAGE									
U	U	U									
			19a. NAME OF RESPONSIBLE PERSON NAVSUBMEDRSCHLAB Commanding Officer								
			19b. TELEPHONE NUMBER (Include area code) 860-694-3263								

Reset

[THIS PAGE INTENTIONALLY LEFT BLANK]

Elucidating Visual Search Behavior and Resulting Performance in Low Prevalence Dynamic Search

Authors:

^{1,2}Krystina Diaz, BA[†]

^{1,2}Margaret Wise, BA[†]

^{1,2}Sylvia Guillory, PhD

¹Jeffrey Bolkhovsky, PhD

²Chad Peltier, PhD

[†]Krystina Diaz and Margaret Wise contributed equally to this work.

¹Naval Submarine Medical Research Laboratory

²Leidos, Inc.

Approved and Released by:

JAMERSON.MATT | Digitally signed by
HEW.HUNTER.126 | JAMERSON.MATTHEW.HUNTE
9308288 | R.1269308288
Date: 2022.10.31 15:45:25 -04'00'

M. H. JAMERSON, CAPT, MSC, USN
Commanding Officer

Naval Submarine Medical Research Laboratory
Submarine Base New London, Box 900
Groton, CT 06349-5900

ADMINISTRATIVE INFORMATION

The views expressed in this report reflect the results of research conducted by the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. This work was supported by funding work unit number F2001. The study protocol was approved by the Naval Submarine Medical Research Laboratory Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects, protocol number NSMRL.2021.0021. I am a military service member or employee of the U.S. Government. This work was prepared as part of my official duties. Title 17, U.S.C., §105 provides that copyright protection under this title is not available for any work of the U.S. Government. Title 17, U.S.C., §101 defines a U.S. Government work as work prepared by a military Service member or employee of the U.S. Government as part of that person's official duties.

Approved for public release: distribution unlimited.

[THIS PAGE INTENTIONALLY LEFT BLANK]

Abstract

Visual search is an essential element of various professions, such as airport security screening, radiology, and sonar monitoring, in which accurate and efficient task performance is vital. Visual search performance is dependent on specific search conditions, such as the frequency of targets, if targets are moving or stationary, and the duration of target visibility. Despite existing research on how conditions such as these individually influence search performance, the extent to which they interact—particularly target prevalence in different types of search environments—is unclear. Therefore, this study investigated performance and perceived cognitive demands of three common aspects of visual search: target prevalence (50% vs 10%), display type (static still-image vs. dynamic continuous scroll), and display time (3.5 vs. 7 seconds). Participants ($n = 556$) recruited through Amazon Mechanical Turk were randomly assigned to one of eight experimental conditions created from the combinations of these search conditions. Participants were tasked with detecting known targets, “T”s, among a display of distractors, offset “L”s, and subsequently rating their perceived condition-specific cognitive task workload via the NASA Task Load Index (NASA-TLX). Though no significant interactions between search conditions on accuracy were found, condition-specific effects were present. Lower target prevalence produced decreased hit rates, replicating the Low Prevalence Effect (LPE) (Wolfe et al., 2007), across static and dynamic displays. Additionally, NASA-TLX ratings implies observer blindness to actual performance decrements in low target prevalence searches. The persistence of the LPE highlights the need for LPE-mitigation methods in both types of search environments.

Keywords: visual search, low prevalence effect, dynamic search, simulated targets, search strategy, cognitive workload

Author Note: Correspondence should be addressed to Jeffrey Bolkhovsky, PhD, Naval Submarine Medical Research Laboratory, Submarine Base New London, Groton, CT 06349-5900, email: jeffrey.b.bolkhovsky.civ@health.mil

Table of Contents

Abstract	6
Introduction.....	9
Methods.....	13
Participants	13
Design.....	13
Stimuli	14
Procedure.....	14
Analysis	15
Results.....	16
Average false alarm rate.....	17
Average hit rate	17
Average reaction time	17
Reaction time variability	18
NASA-TLX Ratings.....	22
Discussion	24
Conclusion	28
Limitations	28
Future Work	29
References.....	30
Appendix A: NASA Task Load Index Rating Scale Definitions	35

Table of Figures

FIGURE 1. A TASK SCHEMATIC FOR (A) STATIC AND (B) DYNAMIC SEARCH DISPLAYS.....	15
FIGURE 2. OBJECTIVE BEHAVIORAL RESULTS.....	20
FIGURE 3. SUBJECTIVE SCORE OF AVERAGE NASA-TLX RATINGS.....	23

Table of Tables

TABLE 1. PARTICIPANT DISTRIBUTION ACROSS EACH TASK CONDITION	14
TABLE 2. MEANS (AND STANDARD ERRORS) FOR PERFORMANCE RESULTS	17
TABLE 3. 3-WAY ANOVAS OF FALSE ALARMS, HIT RATES, REACTION TIMES.....	20
TABLE 4. MEANS (AND STANDARD ERRORS) FOR NASA-TLX RATINGS	22
TABLE 5. 3-WAY ANOVAS OF NASA-TLX RATINGS	23

Introduction

Visual search is an integral component of many professions, with accurate target detection proving vital, e.g., detecting an abnormal growth on a mammogram or identifying a threat in a security surveillance video. Real-world visual search and target detection performance can vary based on multiple task-dependent factors. For example, professionals such as airport security screeners and radiologists examine still-image, or “static” displays (e.g., scans of luggage or MRIs) for low prevalence “target” items (e.g., weapons or tumors, respectively) among more common high prevalence “non-target” items (e.g., socks or normal tissue, respectively). The objects these professionals must identify are often stationary and the amount of time the observers have to perform the search (with or without pause) is not entirely within their control. Airport security screeners, for example, must contend with the fluctuating volume of airport luggage traffic, and may have few opportunities to take a break and pause baggage search. Other occupations, such as submarine sonar monitors, often must search for low-prevalence targets (e.g., foreign submarines among marine life or merchant ships) under conditions where target are moving, or “dynamic”. These operators examine search environments where the search display items may be in constant motion underwater. Due to item motion, targets can appear onscreen and move offscreen at any time. Additionally, due to the real-time nature of these searches, there is no opportunity to quit or pause the search. Therefore, there is limited time to examine every search item that appears onscreen, potentially increasing the likelihood of target misses. All of these condition-dependent search factors can influence operator performance. While often investigated independently, the interaction effects between target prevalence (Horowitz, 2017; Wolfe et al., 2007), display type (Wang & Sun, 2015), and display time (Rieger, Heilmann, & Manzey, 2021; Yu, Yang, Guo, & Zhang, 2015) often go unexplored. Given that real-world visual searches such as sonar monitoring—which holds particular relevance to military research goals to assess sonar technician performance—comprise different combinations of these conditions, further exploration is warranted concerning their individual effects, interactions, and relationships to target detection performance. Understanding the influence of target prevalence under these different conditions can provide key insights into the search behaviors behind the performance effects of speed and display type.

Target Prevalence

Target prevalence and its relationship with visual search performance has been studied extensively. Many studies have shown that a low incidence of targets results in poorer target detection in visual search tasks—a phenomenon known as the “low prevalence effect” (LPE) (Hout, Walenchok, Goldinger, & Wolfe, 2015; Wolfe, Horowitz, & Kenner, 2005). In static visual search tasks, targets are missed two to five times more when target prevalence is 10% compared to a prevalence of 50% (Ishibashi, Kita, & Wolfe, 2012). These misses are typically accompanied by decreased false alarm rates and/or decreased target absent reaction times (RTs) when compared to higher target prevalence visual search tasks (Godwin, Menneer, Riggs, Cave, & Donnelly, 2015; Peltier & Becker, 2016; Schwark, Sandry, Macdonald, & Dologov, 2012; Van Wert, Horowitz, & Wolfe, 2009; Wolfe et al., 2007). This decreased target absent reaction time (the time for the observer to decide that no target is present then make a target absent response) is due to a lowered quitting threshold (the number of items the observer is willing to search before giving up and terminating the search) and a less complete search under low target prevalence relative to high target prevalence. Because LPE is so prevalent across real-world

visual searches and can be devastating in many scenarios (e.g., failure to identify a foreign vessel can result in submariner casualties and/or mission failure), it is crucial to investigate search behaviors and decrements in performance.

Display Type

The search environment can have a profound impact on performance. A static search involves items that are displayed as still images or remain in a constant position and a dynamic search involves items that are moving or changing locations on the screen. Together, these properties define the display type, or the method by which search items are displayed on screen. Studies have found that observers inspecting dynamic displays where both target and non-target objects are moving (unlike predator-prey situations where a moving target against a mostly static background facilitates target detection creating a pop-out effect (Abrams & Christ, 2003)); had poorer target identification performance than observers inspecting static displays (Beck, Peterson, & Vomela, 2006; Stothart, Clement, & Brockmole, 2018). Display type search behaviors and performance differences have often been attributed to observers being able to track fewer dynamic objects accurately due to limited mental resources (Pylyshyn & Storm, 1988; Yantis, 1992). Additional explanations are more subjective, such as increased cognitive workload likely associated with observers adopting different search strategies (Allen, McGeorge, Pearson, & Milne, 2004; Allen, McGeorge, Pearson, & Milne, 2006; Beck, Peterson, & Vomela, 2006; Tombu & Seiffert, 2008). For example, in dynamic searches, which prevent the use of inhibition of return (IOR; (Posner, Rafel, Choate, & Vaughn, 1985); the memory for previously inspected areas/items that facilitates search to new areas/items), search is random, whereas in static displays, search is more systematic with fewer item re-inspections (Wang & Sun, 2015). A random search strategy lacks a fixation pattern leading to items that are repeatedly fixated or never fixated because observers do not have sufficient cognitive resources to track each of the many moving stimuli (Pylyshyn & Storm, 1988; Yantis, 1992). In contrast, in a systematic search strategy there is an eye-gaze pattern which may minimize items being repeatedly fixated. Given the real-world implications that could arise from the adoption of suboptimal visual search strategies in occupations that require examining dynamic displays (e.g., sonar monitoring), exploration of search performance and behavior in different display types under low target prevalence is crucial.

Display Time

Search performance is also degraded by the effects of display time, i.e., how long visual search items are displayed onscreen and/or how quickly the items move within the screen. Previous literature has found that increased time pressure (having less time to inspect and process display items) yielded worse target detection performance. Conversely, increased exposure time to items in static displays had a significant positive effect on inspection accuracy (Rieger, Heilmann, & Manzey, 2021). Moreover, items presented with shorter durations in dynamic display types resulted in worse search performance when compared to prolonged static searches (Rieger, Heilmann, & Manzey, 2021; Rizzi, Buck, & Anderson, 1979; Wang, Zhang, & Klein, 2010; Wentworth & Buck, 1982; Williams & Borow, 1963). Similar to display type effects, mechanisms behind display time effects and time pressure are often posited to be a result of increased cognitive workload (Allen, McGeorge, Pearson, & Milne, 2006; Beck, Peterson, & Vomela, 2006; Hendy, Liao, & Milgram, 1997). Display time, in combination with display type,

has implications upon search behavior and the decision-making process behind low target prevalence search performance outcomes.

Multiple Decision Model

The multiple-decision model (MDM) of Wolfe and Van Wert (2010) is a framework that describes search behavior and performance as functions of target prevalence. According to the model, during visual search an observer makes a decision (target/not a target) about each search item they examine by considering the information accrued from the search display (Wolfe et al., 2007). The search ends once the observer has made a target-present response, or they have accrued enough information to reach a quitting threshold without finding a target—resulting in a self-terminating, target-absent response (i.e., responding for that trial in the negative as opposed to the affirmative to confirm the presence of a target). Furthermore, the model posits that predictable changes in hit rate, false alarm rate, and reaction time from changes in the target prevalence causes shifts in the decision-making criteria (Green & Swets, 1966; Ratcliff & Rouder, 1998), and in the search quitting thresholds (Hout, Walenchok, Goldinger, & Wolfe, 2015; Wolfe & Van Wert, 2010). A shift toward lower quitting thresholds results in decreased target-absent reaction times and higher miss rates (Schwark, Sandry, Macdonald, & Dologov, 2012; Wolfe et al., 2007), suggesting that observers in low target prevalent searches are inspecting fewer display items before deciding that no target is present and thus prematurely quitting search (Peltier & Becker, 2016). Fewer false alarms and increased miss rates (Peltier & Becker, 2016; Wolfe & Van Wert, 2010) are the consequences of a decision criteria shift which involves whether an observer recognizes an inspected item as a target (Green & Swets, 1966; Ratcliff & Rouder, 1998). Both are indicators that observers are classifying display items as non-targets, or distractors, due to more conservative target-qualifying criteria (Peltier & Becker, 2016; Rich et al., 2008; Wolfe & Van Wert, 2010).

The MDM's utility is in using objective performance measures to predict search behavior and performance in self-terminating, static visual search tasks with varying levels of target prevalence. However, the model does not necessarily account for or explain the search behaviors of other real-world search conditions. For example, in some dynamic searches, items are moving throughout the screen and there is a limited and set amount of time that an observer can examine the display items before the items move out of view. There is no longer an option for the observer to manually terminate the search trial and initiate a new trial of search items, thus the quitting threshold portion of the MDM is no longer a factor in this type of search. This circumstance forces the observer to alter their search behavior and make target/not-a-target decisions on as many items as they can within the time-constraints. Investigating conditions outside the original intent of the MDM can help extend the models predictions. As an initial step towards constructing a more comprehensive model of visual search performance, the condition-specific effects of display speed and type must be established.

Relevance of LPE and the MDM to Display Type and Display Time

In non-self-terminating searches, display time and display type limit available search time. The opportunity to make a target-absent response to terminate search, an indicator of reaching a quitting threshold, is not available to measure. In target prevalence manipulations where search time is equated between high and low prevalence searches, it is predicted that target misses would instead be attributable to MDM decision criteria shifts, for which smaller LPEs would be expected. Peltier and Becker (2016) showed that shifts in the quitting threshold

of the MDM account for most of the changes in static visual search performance at different prevalence rates. Early quitting of search accounted for approximately 75% of misses under low prevalence, whereas shifts toward a conservative decision-making criterion accounted for the remaining 25%. Similarly, early quitting accounted for 85% of high prevalence misses and conservative decision-making accounted for the remaining 15%. In a varying target prevalence manipulation where search time is controlled, however, it is expected that the number of misses due to an early quitting threshold would be similar between conditions. We would still expect, for the low prevalence conditions to result in more misses due to the decision criterion shift. These conservative decision criteria-based misses are relatively infrequent, accounting for only 15-25% of errors (Peltier and Becker, 2016). We predict that a time-controlled search task in a low target prevalence condition will result in an increase in misses of approximately 10% compared to the high target prevalence condition.

Because real-world searches (e.g., sonar monitoring), are often time-limited due to the presentation of dynamic, continuously moving stimuli, prematurely quitting the search is often not an option. The interaction among target prevalence, display type, and search time poses a real-world concern. It is necessary to systematically manipulate and assess the aforementioned search conditions on target detection performance to understand their effects on real world visual search outcomes and behavior.

Subjective Performance Measures

The majority of the visual search studies focus on objective measures (e.g., hit rates, false alarm rates, and reaction times); however subjective reports (e.g., self-rated task workload, self-rated performance, etc.) can provide insights into perceived performance or effort. Use of subjective measures in performance-based tasks has shown discrimination between task difficulty where objective performance did not (Hoang, Ranchet, Derollepot, Moreau, & Paire-Ficout, 2020) and provided associations between components of executive control and behavior that objective tools did not elicit (Nelson et al., 2020). Furthermore, Mansikka, Virtanen, and Harris (2019) found that perceived cognitive workload scores were able to distinguish between high, medium, and low levels of performance in a task with manipulated temporal demands, highlighting the ability of subjective reports to serve as an indirect measure of task performance behavior. The use of subjective reports to supplement objective findings can potentially capture additional information regarding search behaviors and strategies.

The Current Study

The current study aims to determine the effects of real-world conditions of visual search on performance, specifically target identification rates (hits), error rates (false alarms), reaction times (RT) and RT variability, juxtaposed with perceived workload. Using hit rates and false alarms, signal detection measures of perceptual sensitivity and response bias were calculated. Perceptual sensitivity and response bias are often used to model discrimination decisions between the signal (target) and noise (distractors). The observed changes in bias criterion with changing prevalence while sensitivity does not systematically vary constitute the basis for the MDM (Wolfe & Van Wert, 2010). We also use RT and RT variability as additional measures of performance, where RT serves as a standard index of task difficulty (faster RTs correspond to easier tasks) and RT variability corresponds to search consistency (high variability is more likely under a random search model, low variability is a more systematic search). We designed a visual search task that utilized components of a sonar monitoring task in which visual search items in a

dynamic task move continuously and horizontally across the screen. This type of visual search task has been relatively unexplored under different prevalence manipulations, as most research has investigated the effects of target prevalence in static searches (see Wolfe, Horowitz, & Kenner, 2005) (see or has not compared performance between static and dynamic searches (see (Mitroff & Biggs, 2014)). We hypothesize that low target prevalence, dynamic displays, and short display times will all yield lower target identification accuracy, increased performance errors, and will be perceived as more demanding.

Methods

Participants

Six hundred and sixty-three participants were recruited through Amazon Mechanical Turk (Amazon MTurk). One-hundred and seven participants were excluded from further analyses for failing to meet the required performance criteria (exclusion criteria detailed in the analysis section), resulting in a final sample size of five hundred and fifty-six.

Design

We tested the interactions of three visual search conditions: target prevalence, display type, and display times. These three factors were manipulated orthogonally in a 2 (target prevalence: high, low) $\times$ 2 (display type: static, dynamic) $\times$ 2 (display time: fast, slow) experimental design, creating eight possible task conditions. The participants were randomly assigned to one of the eight task conditions (Table 1) in a between-subjects design. A minimum required sample size was calculated with the eight groups to approximate an interaction effect of the three conditions using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) for a fixed effect, special, main effects and interactions ANOVA ($\alpha = 0.05$; power $(1-\beta) = 0.90$, f (small-medium effect size) = 0.15). This analysis yielded 469 as the minimum total sample size.

Target prevalence indicated the frequency at which targets appeared throughout the task, which was either high (50% prevalence) or low (10% prevalence). Display type indicated the method by which targets were presented onscreen, which was either a static (still image) or a dynamic (horizontal scrolling) display (Figure 2). Display time indicated how long items appeared on the display or how quickly items moved across the display. The duration was either fast (3.5 seconds) or slow (7 seconds), resulting in a task of approximately 15 minutes or 30 minutes, respectively. With respect to the prevalence condition, an equal number target-present and target-absent trials were maintained across the speed conditions instead of increasing the number of trials in the fast condition to match the overall task duration of the slow condition. A slow condition of 7 seconds was within established reaction time limits for this search array size in a self-terminating search (Peltier & Becker, 2017). In halving the time limit in the fast condition this matched with previous studies manipulating time pressure during search (Rieger, Heilmann, & Manzey, 2021). It should be noted that manipulating the display type, the dynamic motion condition introduces the component of speed. In the static condition the speed would be 0 pixels/s while in the dynamic it would either be approximately 293 pixels/s or 146 pixels/s for the 3.5 and 7 second conditions, respectively. Prior studies have compared static displays to items moving at different velocities while controlling for duration (Hulleman, 2009), however the continuous nature and movement direction of the dynamic condition connected the characteristics of duration with speed. An interaction effect between display type and duration might reveal the impact that item speed has on search by teasing apart difference beyond the

main effect of duration, where the Static Slow/Dynamic Slow and Static Fast/Dynamic Fast have the same duration (7s and 3.5s, respectively).

Table 1. Participant distribution across each task condition

Condition Name	Display Type	Display Time	Target Prevalence	N (collected)	N (analyzed)	Percentage of Participants Retained
DfHP	Dynamic	Fast	High	81	72	89%
DfLP	Dynamic	Fast	Low	80	63	79%
DsHP	Dynamic	Slow	High	93	83	89%
DsLP	Dynamic	Slow	Low	92	73	79%
SfHP	Static	Fast	High	78	65	83%
SfLP	Static	Fast	Low	73	62	85%
SsHP	Static	Slow	High	74	70	95%
SsLP	Static	Slow	Low	82	68	83%

Stimuli

The task was programmed and presented with Inquisit 6 software (Millisecond Software, LLC). For each trial, an image of 1024×768 pixels 300 dpi resolution comprised the search display. The image consisted of 24 items randomly arranged across the screen (Peltier & Becker, 2016). Target-absent images consisted of 24 distractors, offset “L”s. In target-present trials, items comprised 23 distractors, with a single target, “T” (Figure 1). Individual items measured 0.25×0.25 -inches on a 20-inch monitor and could appear in four different orientations: 0° , 90° , 180° , or 270° rotated from their original, upright positions (Figure 1). Individual item orientation was randomly assigned. There were 450 images in the image set (300 total target-absent images and 150 total target-present images) from which 250 images, or trials, were presented in random order depending on the condition’s target prevalence rate. No image was presented more than once.

In the static conditions, the stimuli were presented as still images, with one image per trial (Figure 1A). The stimuli would automatically change to a new image for each trial, and the timing of this change was dependent on the condition-specific display speed (either 3.5 or 7 seconds). A blank white screen was displayed for 500 ms between each trial. In the dynamic conditions, the images/stimuli would move across the screen horizontally, from left to right, continuously until the end of the task with no visible distinction between trials/images (Figure 1B). Dynamic stimuli would be visible to the participant only within the display window for the condition-specific time allotment. Neither the static nor dynamic condition were self-paced, rather the trial/image would advance following the finite stimulus duration of 3.5 (fast) or 7 (slow) seconds.

Procedure

Participants were recruited and compensated \$5.00 USD through MTurk. Each participant completed their assigned task condition on their personal computers and were asked to complete the task in a quiet location to limit distractions around them. Participants were not made aware of their assigned task condition, but they were informed that the task would take approximately 30 minutes to complete. Participants were instructed to visually search their display and indicate a target-present response by left-clicking on their mouse. If no target was detected, they were instructed to let the trial terminate with no response. After receiving instructions, participants began with 10 practice trials to ensure understanding of the task. Upon

conclusion of the practice trials, participants completed 250 experimental task trials. Following the visual search task, participants completed the NASA Task Load Index (NASA-TLX), in which they rated perceived workload on six different metrics: Mental Demand, Physical Demand, Temporal Demand, Effort, Performance, and Frustration Level (Hart & Staveland, 1988). Participants rated each metric on a scale from 1-10, with 1 indicating low demand and 10 indicating high demand. On the “Performance” metric specifically, the rating was reversed with 1 indicating low-performance success and 10 as high-performance success.

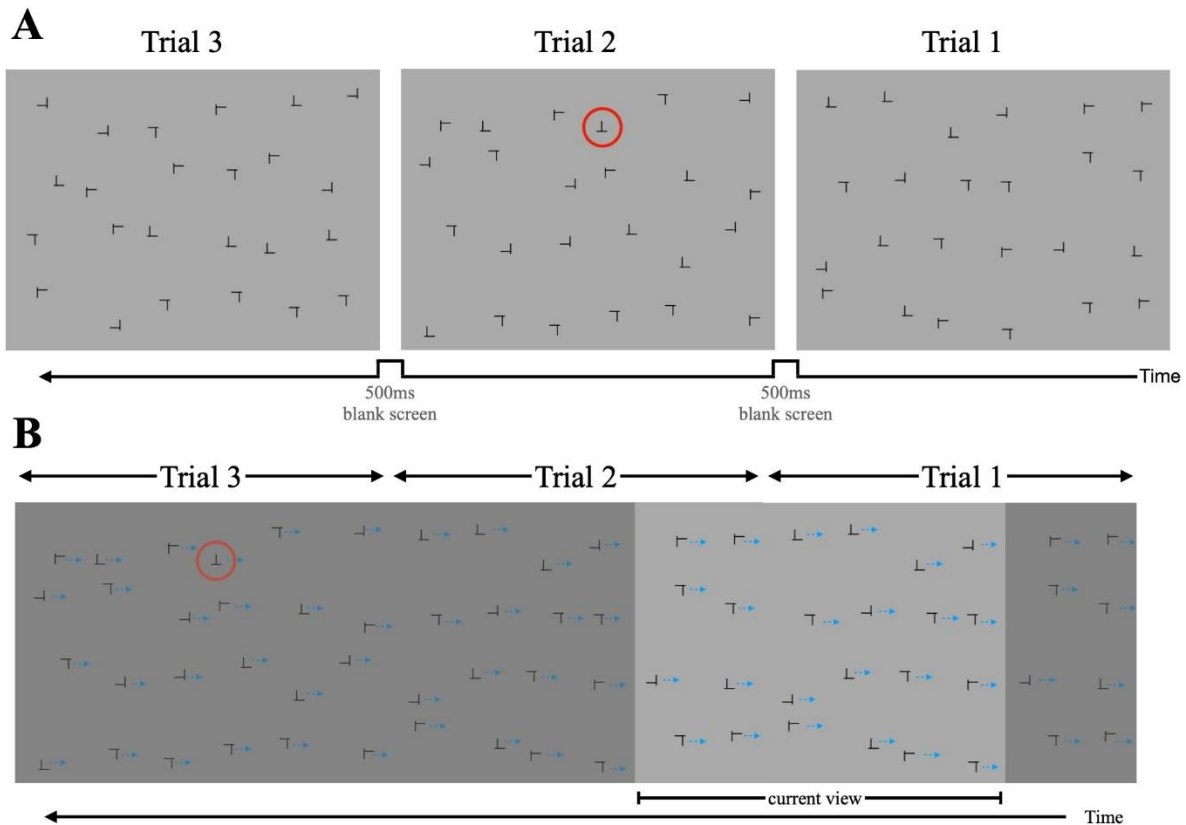


Figure 1. A task schematic for (A) static and (B) dynamic search displays

Participants were instructed to search for a letter “T” (target) among offset “L”s (distractors). The static search display (A) included a 500 ms blank screen between each search display. In the dynamic search display (B), all of the objects moved at the same velocity in a rightward motion (indicated by the dashed blue arrows which were not featured on the actual display). The trials were concatenated together, with no visible break between trials (featured as the highlighted “current view” of the display which is a combination of Trial 1 and Trial 2). In the example only, the target, “T”, is circled in red where those trials are designated as target present trials.

Analysis

Participants were excluded from analysis for failing to complete the task, i.e., completing less than 250 trials or for producing below acceptable performance. The performance criteria involved either the lack of a single response (target-present) throughout the entirety of the task ($n = 33$) or having a false alarm rate greater than or equal to 0.5 which was implemented to exclude excessive responding that corresponded to guessing ($n = 69$). Participants were also excluded due to technical errors in the Inquisit 6 software data output that resulted in an impossible response

time, such that the values were greater than the length of the trial, e.g., 20000+ ms ($n = 5$). The final sample consisted of 556 participants.

The performance measures of interest were: the false alarm rate, hit rate, and the average reaction time of hits (ms). Signal detection measures of non-parametric sensitivity (A') and response bias (B''_D) were also analyzed to quantify the ability of participants to discriminate between targets and distractors and response tendencies (Wickens, 2002).

A trial was considered a false alarm if the participant made a target-present response (mouse click) during a target-absent trial, i.e., when the target item was not visible on the display. Likewise, in a target-present trial, the trial was categorized as a hit if the first target-present response was made while the target item was present on the display; the response was categorized as a miss upon the start of a new trial/as soon as the target disappeared. The reaction time data were calculated using target-present trials and were from the time the target stimulus was visible onscreen to the time the participant made a target-present response.

As a consequence of the linear uniform motion of the dynamic condition, the predictability of the movement might benefit search strategies that promote the allocation of attention to a specific area in the search display. In monitoring for the target to eventually traverse through this limited area, a participant can passively wait for the target instead of actively searching the display. To evaluate for this possibility the standard deviation of reaction time to hits were analyzed to provide insight into observer search strategies (e.g., random search, sit-and-wait search, etc.). In the dynamic condition, a sit-and-wait strategy would result in less variation in the time used to locate the target because they would be finding the target at roughly the same time if the participant maintained their attention on the same area of the display compared to a random search which would exhibit increased variability and is behavior expected during a static search.

Participant NASA-TLX ratings of perceived cognitive workload and difficulty, scored on a 10-point scale, were collected as an exploratory analysis, to measure self-reported task difficulty and sensitivity to performance (for more information, please see Appendices A). For each objective task measure a $2 \times 2 \times 2$ univariate analysis of variance (ANOVA) was used to determine if there were any significant effects of prevalence, display speed, or display type (at $p < 0.05$). Follow-up pairwise comparisons with Bonferroni corrections were used to determine the direction of any significant effects. In instances of a significant interaction effect, follow-up ANOVAs were parsed into lower order ANOVAs. When homogeneity of variance was violated, a ratio of the sample size for the task conditions was calculated to determine whether the conditions were approximately equal in size such that the ratio was below 1.5 (Blanca, Alarcón, Arnau, Bono, & Bendayan, 2018). There were no instances where the ratio fell below the criteria. In analyzing the subjective NASA-TLX ratings, each of the 6 facets of the NASA-TLX (mental demand, physical demand, temporal demand, performance, effort, and frustration) were explored as a function of the three factors of task conditions separately. As these analyses were exploratory, family-wise corrections were not applied, however, corrected p -values are presented for comparison.

Results

Average behavioral performance for each condition is summarized in Table 2 and statistical test results are summarized in Table 3.

Table 2. Means (and Standard Errors) for Performance Results

Search Task Condition	Hit Rate	False Alarm Rate	Average RT (ms)	A'	B''D	RT Variability (ms)
DfHP	0.22 (0.01)	0.07 (0.01)	2043.65 (43.09)	0.69 (0.02)	0.63 (0.04)	665.74 (21.66)
DfLP	0.20 (0.02)	0.05 (0.01)	2039.87 (64.48)	0.66 (0.02)	0.73 (0.04)	613.86 (40.30)
DsHP	0.36 (0.02)	0.06 (0.01)	1569.41 (35.88)	0.75 (0.02)	0.68 (0.04)	709.27 (27.45)
DsLP	0.34 (0.03)	0.06 (0.01)	3370.08 (125.42)	0.70 (0.02)	0.77 (0.04)	1038.08 (72.39)
SfHP	0.24 (0.02)	0.09 (0.02)	1899.34 (55.61)	0.70 (0.02)	0.55 (0.04)	882.66 (22.61)
SfLP	0.21 (0.02)	0.07 (0.01)	2044.72 (70.00)	0.70 (0.02)	0.64 (0.04)	708.42 (40.68)
SsHP	0.39 (0.02)	0.06 (0.01)	3683.88 (87.29)	0.79 (0.02)	0.68 (0.04)	1668.95 (32.01)
SsLP	0.34 (0.02)	0.05 (0.01)	3805.75 (125.99)	0.75 (0.02)	0.78 (0.04)	1555.68 (55.01)
Fast	0.22 (0.01)	0.07 (0.01)	2006.90 (41.43)	0.69 (0.01)	0.64 (0.02)	717.67 (21.04)
Slow	0.36 (0.01)	0.05 (0.01)	3107.28 (39.29)	0.75 (0.01)	0.73 (0.02)	1242.99 (19.81)
Static	0.29 (0.01)	0.07 (0.01)	2858.43 (40.62)	0.74 (0.01)	0.66 (0.02)	1203.93 (20.53)
Dynamic	0.28 (0.01)	0.06 (0.01)	2255.75 (40.13)	0.70 (0.01)	0.70 (0.02)	756.74 (20.34)
High Prevalence	0.30 (0.01)	0.07 (0.01)	2299.07 (38.53)	0.73 (0.01)	0.63 (0.02)	981.65 (19.33)
Low Prevalence	0.28 (0.01)	0.06 (0.01)	2815.11 (42.14)	0.71 (0.01)	0.73 (0.02)	979.01 (21.49)

Average false alarm rate

Statistical analyses did not reveal significant main effects of visual search condition on false alarm rate ($p > .05$). Moreover, no significant interaction effects were detected (Table 2, Table 3). No measures of effect size exceeded Cohen's $d = .07$.

Average hit rate

A main effect of prevalence was found to be statistically significant, ($p = .013$) with pairwise comparisons revealing that low prevalence searches resulted in a lower hit rate compared with high prevalence searches. Significant differences in display times were also detected ($p < .001$), in which the fast conditions yielded significantly lower hit rates than the slow conditions (display time variables are defined in Table 1 and page 15). No significant effects were found between the static or dynamic display types, ($p = .29$) and no significant interaction effects were found ($p > .05$). Condition and group effects can be seen in Figure 2A and Figure 2B, respectively. Effect sizes showed a small (Cohen's $d = .11$) gain in the average hit rate for high target prevalence relative to low and in slow display time compared to fast (Cohen's $d = .43$).

Average reaction time

Results from the average reaction time (ms) of target hits revealed significant differences between conditions (Figure 2C and Figure 2D, Table 2). Prevalence conditions were significantly different, ($p = .001$), such that that low prevalence conditions had longer response latencies than high prevalence conditions. A significant main effect of Display Time was also found, ($p = .001$) and a large effect size (Cohen's $d = .84$) where slow display times resulted in greater average reaction time latencies than fast. Display type also yielded a significant main effect, ($p = .001$) for which target-present reaction times on static displays took a greater amount of time than in dynamic displays (it should be noted that this result is not in any way surprising: slow conditions' max RTs are 7,000ms, whereas fast conditions' max RTs are 3,500ms). There was a significant two-way interaction between display time and display type, ($p = .001$); time and

prevalence ($p = .001$); and display type and prevalence, ($p = .001$). The effect size for the interaction between Display Time and Type indicated a slowing in reaction times within the Slow Static search condition relative to the other conditions (Cohen's $d = .51$). There was also a significant three-way interaction between and display time, display type, and prevalence ($p = .001$).

A follow-up two way ANOVA of the three-way interaction between display type and target prevalence in the fast condition found no significant effects (main effect of target prevalence: $F(1, 247) = 1.44, p = .23, \eta_p^2 = 0.01$; main effect of display type: $F(1, 247) = 1.49, p = .22, \eta_p^2 = 0.01$; target prevalence $\times$ display type: $F(1, 247) = 1.65, p = .20, \eta_p^2 = 0.01$). There were significant main effects of display type ($F(1, 275) = 102.65, p < .001, \eta_p^2 = 0.27$) and prevalence in the slow condition ($F(1, 275) = 180.60, p < .001, \eta_p^2 = 0.40$) as well as a significant interaction ($F(1, 275) = 78.27, p < .001, \eta_p^2 = 0.22$). Low prevalence search (3587.91 ± 69.93 ms) yielded approximately a 37% greater mean reaction time than high prevalence (2626.65 ± 64.12 ms) in the slow condition. Additionally, the static display type (3744.82 ± 68.15 ms) resulted in longer reaction times than the dynamic display (2469.75 ± 66.02 ms) in slow task conditions. The slowing of reaction times between the low to high target prevalence was more pronounced in the dynamic condition (low: 3370.08 ± 99.29 ms; high: 1569.41 ± 87.03 ms) than for static (low: 3805.75 ± 98.51 ms; high: 3686.88 ± 94.19 ms) within the slow display time condition.

Signal Detection Measures (A' and B''_D)

The sensitivity and response bias measures were calculated using the R software package Psycho (Makowski, 2018). Results are summarized in Table 2 and Table 3.

The analysis for sensitivity revealed a main effect of presentation rate ($p < .001$), such that the slow condition had an increased sensitivity compared to the fast condition. A main effect of display type ($p = .013$) found that the static condition had an increased sensitivity in comparison to the dynamic condition. The effect of prevalence was not significant ($p = .55$) and none of the interactions reached statistical significance.

The response bias analysis found a conservative bias (increase in no-signal responses) in the slow relative to the fast condition ($p = .002$). There was a significant bias in target prevalence ($p = .001$), where the low prevalence condition had a higher bias towards the no-signal response than the high prevalence condition. No other main effects or interactions were significant.

In summary, participant sensitivity (A') for discriminating targets from distractor objects was higher when the objects were not moving (static condition) compared to when the objects were dynamic and when there was more time to inspect objects (slow). Furthermore, there was increased conservatism with respect to responding in the slow compared to fast, and in the low compared to high target prevalence.

Reaction time variability

To investigate the possible use of multiple search strategies, the variability of the reaction time (standard deviation) data were submitted to an ANOVA. Low prevalence conditions were not found to significantly differ from high prevalence conditions ($p = .93$) (Figure 2E and Figure 2F). Slow display times had significantly greater variability than fast ($p < .001$; Cohen's $d = .81$). The analyses also revealed a significant difference in display type ($p < .001$) and a medium effect size (Cohen's $d = .69$) with static displays having greater response time variability than dynamic displays. Significant interaction effects were found between

display time and type ($p < .001$); display time and prevalence ($p < .001$); type and prevalence ($p < .001$); and display time and type and prevalence ($p = .006$).

Follow-up two-way ANOVAs to parse the three-way interaction into lower order ANOVAs by display time showed that there was a significant difference in target prevalence ($F(1, 237) = 13.15, p < .001, \eta_p^2 = 0.05$) and in display type ($F(1, 237) = 24.96, p < .001, \eta_p^2 = 0.10$) in the fast conditions. With regards to prevalence, low yielded less variability (661.14 ± 20.87) than high (774.20 ± 23.16) in fast conditions, and static tasks had greater variability (795.54 ± 21.96) than dynamic tasks (639.80 ± 22.12) in fast conditions. In slow conditions, target prevalence ($F(1, 268) = 5.29, p = .02, \eta_p^2 = 0.02$) and display type ($F(1, 268) = 248.33, p < .001, \eta_p^2 = 0.48$) bore significant differences, in which low prevalence conditions had greater variability (1296.88 ± 34.88) than high prevalence conditions (1189.11 ± 31.31), and static conditions had almost twice the variability (1612.31 ± 33.60) of the dynamic conditions (873.68 ± 32.68).

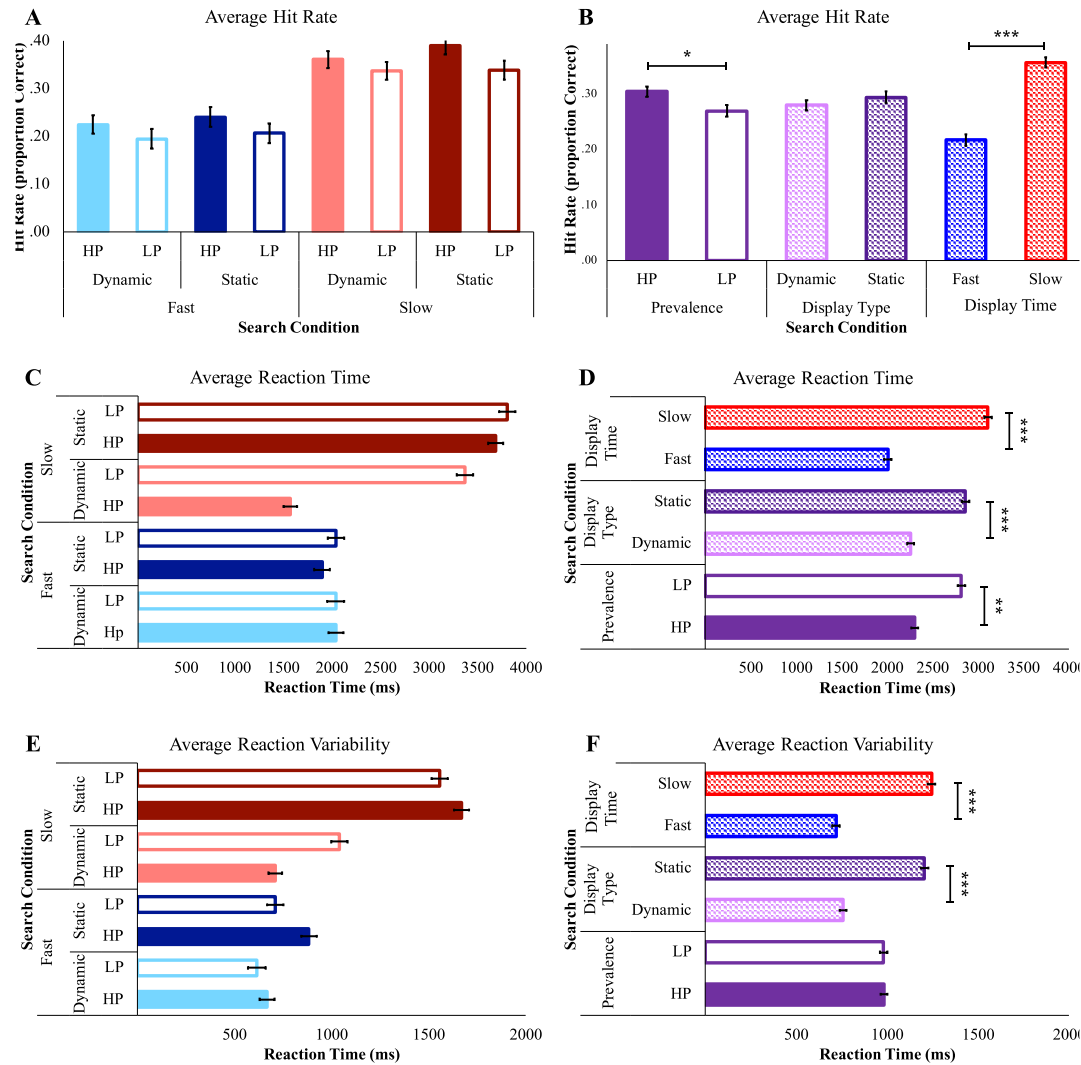


Figure 2. Objective Behavioral Results

(A) The mean hit rates of each task condition across main effects where solid bars represent the high prevalence (HP) condition and unfilled bars the low prevalence condition (LP), dark colors indicate the static condition and light colors the dynamic, and where the red colors correspond to the slow conditions and blues the fast condition. (B) The combined mean hit rates across all condition levels (main effects). (C) The mean reaction times of each task condition across main effects. (D) The combined mean reaction times across all condition levels (main effects). (E) The reaction time variability of each task condition across main effects. (F) The averaged reaction time variability (standard deviations) across all condition levels (main effects). Error bars ± 1 standard error (SE). Significant main effects at $p < 0.05$ denoted with an “*”, $p < .01$ as “**”, and $p < .001$ as “***”.

Table 3. 3-Way ANOVAs of False Alarms, Hit Rates, Reaction Times

		F value (df)	<i>p</i> -value	η_p^2	Cohen's <i>d</i>
false alarms	Prevalence	2.12 (1,548)	.15	0.004	0.06
	Display Type	0.58 (1,548)	.45	0.001	0.03
	Display Time	2.46 (1,548)	.15	0.004	0.07

	Prevalence $\times$ Type	0.07 (1,548)	.80	<0.001	0.01
	Time $\times$ Type	1.45 (1,548)	.23	0.003	0.05
	Time $\times$ Prevalence	1.25 (1,548)	.27	0.002	0.05
	Time $\times$ Type $\times$ Prevalence	0.12 (1,548)	.73	<0.001	0.01
average	Prevalence	6.23 (1,548)	.01*	0.01	0.11
hit	Display Type	1.11 (1,548)	.29	0.02	0.04
rate	Display Time	100.90 (1,548)	<.001*	0.16	0.43 [#]
	Prevalence $\times$ Type	0.35 (1,548)	.55	0.001	0.03
	Time $\times$ Type	.004 (1,548)	.95	<0.001	0
	Time $\times$ Prevalence	0.05 (1,548)	.83	<0.001	0.01
	Time $\times$ Type $\times$ Prevalence	0.19 (1,548)	.66	<0.001	0.02
reaction	Prevalence	81.68 (1,522)	.001*	0.14	0.4 [#]
time	Display Type	111.41 (1,522)	.001*	0.18	0.46 [#]
average(ms)	Display Time	371.40 (1,522)	.001*	0.42	0.84 ^{###}
	Prevalence $\times$ Type	44.86 (1,522)	.001*	0.08	0.29 [#]
	Time $\times$ Type	138.68 (1,522)	.001*	0.21	0.51 ^{##}
	Time $\times$ Prevalence	60.80 (1,522)	.001*	0.10	0.34 [#]
	Time $\times$ Type $\times$ Prevalence	64.06 (1,522)	.001*	0.11	0.35 [#]
A'	Prevalence	3.71 (1,548)	.55	0.007	0.08
	Display Type	6.20 (1,548)	.01*	0.011	0.11
	Display Time	16.00 (1,548)	<.001*	0.028	0.17
	Prevalence $\times$ Type	.71 (1,548)	.40	0.001	0.04
	Time $\times$ Type	.40 (1,548)	.53	0.001	0.03
	Time $\times$ Prevalence	.63 (1,548)	.43	0.001	0.03
	Time $\times$ Type $\times$ Prevalence	.14 (1,548)	.71	<0.001	0.02
B'' _D	Prevalence	12.15 (1,548)	.001*	0.022	0.15
	Display Type	1.96 (1,548)	.16	0.004	0.06
	Display Time	9.68 (1,548)	.002*	0.017	0.13
	Prevalence $\times$ Type	0.01 (1,548)	.92	<0.001	0
	Time $\times$ Type	2.54 (1,548)	.11	0.005	0.07
	Time $\times$ Prevalence	0.002 (1,548)	.97	<0.001	0
	Time $\times$ Type $\times$ Prevalence	0.07 (1,548)	.79	<0.001	0.01
reaction	Prevalence	0.01 (1,505)	.93	<0.001	0
time	Display Type	239..41 (1,505)	<.001*	0.32	0.69 ^{##}
variability	Display Time	330 (1,505)	<.001*	0.40	0.81 ^{###}
(SD)	Prevalence $\times$ Type	23.83 (1,505)	<.001*	0.05	0.22 [#]
	Time $\times$ Type	101.69 (1,505)	<.001*	0.17	0.45 [#]
	Time $\times$ Prevalence	14.60 (1,505)	<.001*	0.03	0.17
	Time $\times$ Type $\times$ Prevalence	7.65 (1,505)	.006*	0.02	0.12

Significant effects are denoted with an “*” at $p < .05$. Effect sizes (Cohen’s d) are denoted with an “#” at $d > .2$ (small), “##” at $d > .5$ (medium), and “###” at $d > .8$.

NASA-TLX Ratings

Responses to each item (six items in total) on the NASA-TLX questionnaire were analyzed with a between-subjects 2×2×2 ANOVA (Table 3, Table 4). The condition effects are shown in Figure 3. A significant main effect of display time was found on Temporal Demand ($p < .0001$), $\eta_p^2 = 0.05$; Frustration ($p = .028$); and self-rated Performance ($p = .03$). Pairwise comparisons revealed that, on average, fast displays had higher ratings of Temporal Demand than slow displays. Higher ratings on Frustration were found on the fast condition compared to the slow. In addition, participants gave higher ratings on Performance in fast conditions, indicating a worse perceived performance, when compared to slow conditions.

There was a significant main effect of target prevalence on Physical Demand ($p = .014$) and Frustration ($p = .013$), that showed high target prevalence had increased ratings of Physical Demand when compared to the low prevalence conditions, and low target prevalence resulted in increased ratings of Frustration in comparison to high.

There was a significant main effect of display type on Mental Demand ($p = .012$), Temporal Demand ($p < .001$), Effort ($p = .018$); and Frustration ($p = .001$). Static display types yielded higher ratings of perceived Mental Demand, Temporal Demand, Effort, and Frustration than dynamic display types.

A significant interaction effect was found between display type and display time on Temporal Demand ($p = .023$) and perceived Performance ($p = .021$) in which follow-up ANOVAs by presentation type showed that the static-fast conditions yielded higher ratings of Temporal Demand (7.73 ± 0.23) than static-slow (6.02 ± 0.22 ; $p < .001$), and that the dynamic-fast (6.17 ± 0.22) had higher ratings than dynamic-slow (5.45 ± 0.21 ; $p = .028$) conditions. Likewise, a follow-up ANOVA noted that participants gave higher ratings on Performance in the static-fast condition (6.02 ± 0.23) than in the static-slow (5.04 ± 0.22 ; $p < .001$), but that there was no significant difference in ratings between dynamic-fast (5.54 ± 0.22), and dynamic-slow (5.57 ± 0.21 ; $p = .92$) conditions.

Table 4. Means (and Standard Errors) for NASA-TLX Ratings

Search Condition	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
DfHP	7.58 (0.25)	5.71 (0.35)	5.78 (0.30)	5.26 (0.30)	7.72 (0.24)	5.15 (0.34)
DfLP	7.65 (0.27)	5.00 (0.38)	6.56 (0.32)	5.81 (0.32)	7.76 (0.26)	5.65 (0.36)
DsHP	7.99 (0.24)	5.47 (0.35)	5.47 (0.28)	5.48 (0.28)	7.71 (0.22)	4.67 (0.31)
DsLP	7.39 (0.25)	5.10 (0.36)	5.43 (0.30)	5.65 (0.30)	7.57 (0.24)	5.50 (0.34)
SfHP	8.23 (0.27)	5.35 (0.37)	7.52 (0.32)	6.23 (0.32)	8.37 (0.25)	6.14 (0.35)
SfLP	8.19 (0.27)	5.06 (0.38)	7.94 (0.22)	5.81 (0.32)	8.32 (0.26)	6.66 (0.36)
SsHP	8.29 (0.26)	5.57 (0.36)	6.20 (0.31)	5.26 (0.31)	8.09 (0.24)	5.36 (0.34)
SsLP	7.74 (0.26)	4.44 (0.36)	5.84 (0.31)	4.82 (0.31)	7.65 (0.25)	5.93 (0.35)
Fast	7.92 (0.13)	5.28 (0.19)	6.95 (0.16)	5.78 (0.16)	8.04 (0.13)	5.90 (0.18)
Slow	7.85 (0.13)	5.15 (0.18)	5.74 (0.15)	5.30 (0.15)	7.75 (0.12)	5.37 (0.17)
Static	8.11 (0.13)	5.11 (0.18)	6.87 (0.16)	5.53 (0.16)	8.11 (0.13)	6.02 (0.16)
Dynamic	7.65 (0.13)	5.32 (0.18)	5.81 (0.15)	5.55 (0.15)	7.69 (0.12)	5.25 (0.17)

High Prevalence	8.02 (0.13)	5.53 (0.18)	6.24 (0.15)	5.56 (0.15)	7.97 (0.12)	5.33 (0.17)
Low Prevalence	7.74 (0.13)	4.90 (0.18)	6.44 (0.16)	5.52 (0.16)	7.83 (0.13)	5.94 (0.18)

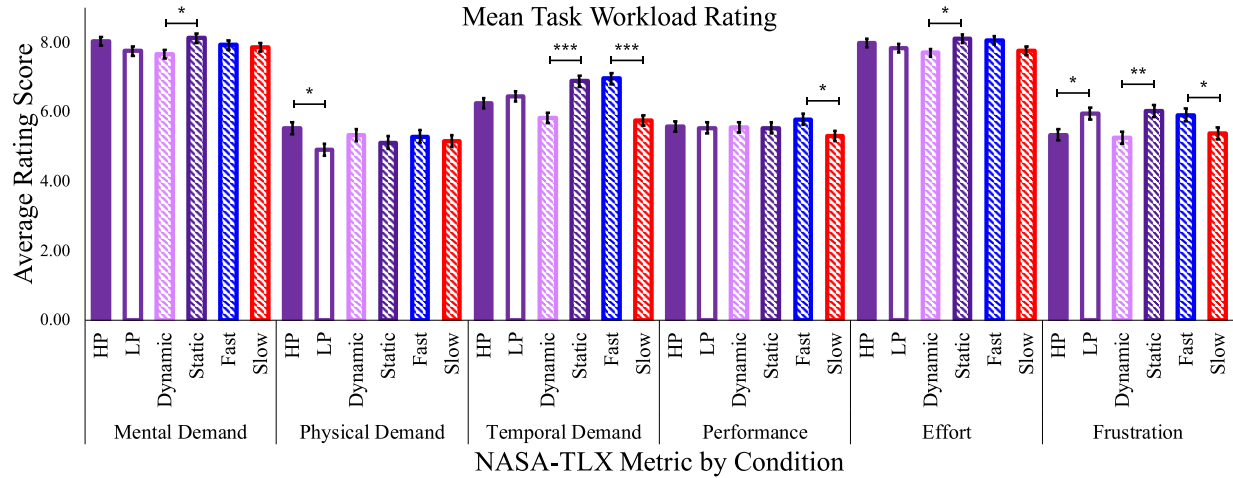


Figure 3. Subjective Score of Average NASA-TLX ratings

The mean NASA-TLX ratings of each of the 6 sub-scales, across each task condition. Error bars ± 1 standard error (SE). Significant effects at $p < 0.05$ denoted with an “*”, $p < .01$ as “**”, and $p < .001$ as “***”.

Table 5. 3-Way ANOVAs of NASA-TLX Ratings

		F value (df)	p-value (uncorrected)	p-value (corrected)	η_p^2	Cohen's <i>d</i>
Mental Demand	Prevalence	2.35 (1,547)	.13	.77	0.004	0.06
	Display Type	6.32 (1,547)	.01*	.07	0.01	0.1
	Display Time	0.13 (1,547)	.72	1.00	< 0.001	<0.032
	Prevalence $\times$ Type	0.13 (1,547)	.94	1.00	< 0.001	<0.032
	Time $\times$ Type	0.13 (1,547)	.46	1.00	0.001	0.03
	Time $\times$ Prevalence	0.13 (1,547)	.11	.64	0.01	0.1
	Time $\times$ Type $\times$ Prevalence	0.13 (1,547)	.83	1.00	< 0.001	<0.032
Physical Demand	Prevalence	6.05 (1,547)	.01*	.08	0.01	0.1
	Display Type	0.69 (1,547)	.41	1.00	0.001	0.03
	Display Time	0.29 (1,547)	.59	1.00	0.001	0.03
	Prevalence $\times$ Type	0.11 (1,547)	.74	1.00	< 0.001	<0.032
	Time $\times$ Type	0.07 (1,547)	.8	1.00	< 0.001	<0.032
	Time $\times$ Prevalence	0.25 (1,547)	.62	1.00	< 0.001	<0.032
	Time $\times$ Type $\times$ Prevalence	1.34 (1,547)	.25	1.00	0.002	0.04
Temporal Demand	Prevalence	0.82 (1,547)	.37	1.00	0.001	0.03
	Display Type	23.82 (1,547)	<.001*	<.001*	0.04	0.2 [#]
	Display Time	30.88 (1,547)	<.001*	<.001*	0.05	0.23 [#]

	Prevalence × Type	0.62 (1,547)	.43	1.00	0.001	0.03
	Time × Type	5.18 (1,547)	.02*	.14	0.001	0.03
	Time × Prevalence	3.32 (1,547)	.07	.41	0.01	0.1
	Time × Type × Prevalence	0.002 (1,547)	.96	1.00	< 0.001	<0.032
Performance	Prevalence	0.03 (1,547)	.87	1.00	< 0.001	<0.032
	Display Type	0.01 (1,547)	.92	1.00	< 0.001	<0.032
	Display Time	4.73 (1,547)	.03*	.18	0.01	0.1
	Prevalence × Type	3.26 (1,547)	.07	.43	0.01	0.1
	Time × Type	5.36 (1,547)	.02*	.13	0.01	0.1
	Time × Prevalence	0.19 (1,547)	.66	1.00	< 0.001	<0.032
	Time × Type × Prevalence	0.18 (1,547)	.68	1.00	< 0.001	<0.032
Effort	Prevalence	0.71 (1,547)	.40	1.00	0.001	0.03
	Display Type	5.66 (1,547)	.02	.11	0.01	0.1
	Display Time	2.78 (1,547)	.10	.58	0.01	0.1
	Prevalence × Type	0.3 (1,547)	.58	1.00	0.001	0.03
	Time × Type	1.17 (1,547)	.28	1.00	0.002	0.04
	Time × Prevalence	0.67 (1,547)	.41	1.00	0.001	0.03
	Time × Type × Prevalence	0.09 (1,547)	.76	1.00	< 0.001	<0.032
Frustration	Prevalence	6.17 (1,547)	.01*	.08	0.01	0.1
	Display Type	10.2 (1,547)	.001*	.01*	0.02	0.14
	Display Time	4.87 (1,547)	.03*	.17	0.01	0.1
	Prevalence × Type	0.06 (1,547)	.81	1.00	< 0.001	<0.032
	Time × Type	0.83 (1,547)	.36	1.00	0.002	0.04
	Time × Prevalence	0.15 (1,547)	.70	1.00	< 0.001	<0.032
	Time × Type × Prevalence	0.08 (1,547)	.77	1.00	< 0.001	<0.032

Significant main effects are denoted with an “*”, at $p < .05$. Effect sizes (Cohen’s d) are denoted with an “#” at $d > .2$ (small), “##” at $d > .5$ (medium), and “###” at $d > .8$.

Discussion

The present study examined the impact that display type and display time under high and low target prevalence had on visual search performance and found results that aligned with and expanded on previous literature. With regards to target identification accuracy, low target prevalence conditions and fast display times yielded greater errors, i.e., reduced hit rates, aligning with the initial hypotheses. The LPE was present across static and dynamic searches and the shift towards a conservative response bias gives further indication of a decision threshold shift. Together, these results are predicted by the MDM and demonstrate the applicability of the model to dynamic and time-restricted visual search performance. The reaction time data provided additional insight into potential search behaviors behind the performance outcomes and the NASA-TLX data too—evidence of the cognitive load participants experienced—supplement conclusions drawn from the more objective performance results.

Target Prevalence

The LPEs observed across conditions align with trends found across previous literature, with observers making significantly fewer hits/greater misses in low target prevalence than high target prevalence conditions (Godwin, Menneer, Riggs, Cave, & Donnelly, 2015; Ishibashi, Kita, & Wolfe, 2012; Rich et al., 2008; Rich et al., 2008; Van Wert, Horowitz, & Wolfe, 2009; Wolfe et al., 2007; Wolfe, Horowitz, & Kenner, 2005). The current study's LPEs showed a reduction in target detection by approximately 13% across all conditions, reflecting the power of MDM decision threshold shifts in the absence of quitting threshold shifts (which were controlled for by the trials' presentation times). These findings are also in agreement with previous literature, in which decision criteria shifts resulted in approximately 10% fewer low-prevalence target detections (Peltier & Becker, 2016). Though the effect of prevalence on hit rate was small (Cohen's $d = .11$), this is likely because quitting thresholds, the major contributor to the LPE, were controlled for by display time, leaving only shifts in decision criteria to cause different rates of miss errors across prevalence rates. Reaction time results revealed increased response latency on hits for low prevalence conditions, which provides additional evidence for the decision criteria theory of the MDM. The current experiment controlled for shifts in the quitting threshold that typically result in reduced response times as a consequence of observers terminating search without inspecting the entire search display (Schwark, Sandry, Macdonald, & Dologov, 2012; Wolfe et al., 2007; Wolfe & Van Wert, 2010). Misses, therefore, could instead be attributed to conservative decision criteria shifts. A conservative shift occurs when observers are less likely to classify a fixated item as a target and results in more time expended accumulating target-qualifying evidence towards a target-present criteria decision boundary. Within the MDM framework the increased reaction times in low prevalence search conditions could reflect a decision criteria shift with the bias, modeled as the starting point for evidence-accumulation, positioned closer to a distractor boundary leading to slow target decisions. These results indicate a more time-consuming and less efficient search strategy during low target prevalence conditions. Moreover, observers may have implemented other decision-making strategies affecting search performance and behavior to accommodate search conditions in which the display time limited item inspection and where the dynamic display presented items in constant motion.

Display Time

With half the amount of time provided for observers to inspect and process the display items, fast display times produced an approximately 65% decrease in target detections, a moderate-sized effect (Cohen's $d = .43$). The fast display times leading to fewer target detections is consistent with previous velocity-based search literature (Erikson, 1964; Wang, Zhang, & Klein, 2010; Williams & Borow, 1963), with some studies positing these decrements to be attributed to improper item formation or capture in which individuals do not have enough time to process the items on the display (Burg, 1966; Wentworth & Buck, 1982; Yantis & Gibson, 1994). Accompanying the reduced hit rates, the faster display times resulted in a small reduction in visual search sensitivity (Cohen's $d = .17$) and slightly biased responses in a more liberal direction (Cohen's $d = .13$). Faster display times typically required less time and had much less variability (Cohen's $d = .81$) in the time it took to identify a target, seeming to align with established literature that has observed decreased reaction times in faster than in slower display time tasks, with observers being less accurate in their faster responses (Rieger, Heilmann, & Manzey, 2021; Yang, Song, Zhang, & Yu, 2016).

A result contrary to the display time manipulations would be surprising because the slow conditions provided more time in which observers could respond, so the reaction time data are naturally higher compared to the fast conditions. It is notable, though, that the reaction time effects observed in the current study's slow display time conditions were largely attributed to low prevalence and static display type task conditions, which resulted in increased response times. This result may be explained by findings from Yu and Chan (2015), which imply that observers employ less efficient search strategies and randomly search the screen (Yu & Chan, 2015). In support of a random search strategy resulting in increased response times, studies documenting inhibition of return (IOR) during visual search, in which fixations on a search item are not repeated, have observed IOR in more well-performed searches (Wang, Zhang, & Klein, 2010); thus, the increased response latencies in the current study may be evidence of decreased IOR, with observers re-fixating items. Performance decrements associated with random search are similar to the ones found in the current study in which items presented for shorter amounts of time or items presented as faster-moving yielded more misses (Yu & Chan, 2015), perhaps an indication of lower perceived workload for low prevalence searches.

Display Type

Contributing to the random search theory are the effects of display type on performance. While the present study did not find significant differences in target identification performance between static and dynamic display type conditions (Cohen's $d = .04$), the reaction time data are potentially more informative. The initial assumption that the dynamic condition's continuous searching and lack of mental respite during item would decrease performance did not account for a random search strategy in the static condition, and a more efficient search strategy employed in a predictable dynamic task. Despite no significant differences detection in mean hit rates as a function of display type, analyses revealed slight differences in dynamic and static sensitivity (Cohen's $d = .11$), moderate effects on reaction times (Cohen's $d = .46$), and a large effect on reaction time variability (Cohen's $d = .69$), signifying those different strategies likely were employed in each condition. The absence of motion increased sensitivity, potentially because participants could choose to focus on a single item to maximally differentiate a target from a potential distractor. Longer average reaction times for hits and increased reaction time variability in static conditions are evidence in support of a random search (Yu & Chan, 2015), with observers haphazardly re-fixating/returning to previously inspected items and being equally likely to find a target on their first and last inspections of a trial; whereas the shorter average reaction times and much less variability in these reaction times in dynamic task conditions allude to a more efficient, more consistent search strategy in which observers monitored only a portion of the screen.

In contrast to previous studies featuring dynamic display conditions with random item-movement (Horowitz & Wolfe, 1998; Peltier & Becker, 2015; Pylyshyn & Storm, 1988) wherein search items enter and/or move throughout the screen in unpredictable and non-uniform directions, the current study employed a continuous horizontal scrolling motion for the dynamic display. All search items entered from the left side of the screen and scrolled at a steady rate across to the right where they disappeared from the display window. The predictable and uniform movement of the continuous scrolling dynamic display items could have enabled faster search (Alvarez, Konkle, & Oliva, 2007; Wentworth & Buck, 1982) and allowed for participants to adopt a 'sit-and-wait' search strategy, which has previously been applied in continuous-type search tasks (Alvarez, Konkle, & Oliva, 2007; Beck, Peterson, & Vomela, 2006; Boot, Kramer,

Becic, Wiegmann, & Kubose, 2006; von Mühlenen, Müller, & Müller, 2003), enabling the observer to focus their attention exclusively to the left side of the display screen in anticipation of the target item entering their visual field. Eye-tracking studies have shown that observers who adopted search strategies with fewer eye movements (e.g., sit-and wait) displayed increased efficiency in search when compared to those who adopted search strategies with increased eye movements (e.g., random search) (Boot, Becic, & Kramer, 2007; Geyer, von Mühlenen, & Müller, 2007). Similarly, participants in a visual search task featuring a dynamic condition with unpredictable, or random item movement displayed worsened efficiency and accuracy (Peltier & Becker, 2015). The predictability of the item movement in the dynamic display that potentially promotes decreased eye-movements could have improved observer search efficiency, yielding search performance rates that were not significantly distinguishable from those in the static condition, despite previous findings (Alvarez, Konkle, & Oliva, 2007). Predictability in item movement has also been linked to decreased cognitive demands of a task (Alvarez, Konkle, & Oliva, 2007; Rieger, Heilmann, & Manzey, 2021; von Mühlenen, Müller, & Müller, 2003). This could explain the NASA-TLX findings in which observers rated the dynamic task conditions as significantly less mentally demanding, temporally demanding, effortful, and frustrating than the static conditions. Despite the decreased perceived cognitive workload associated with the dynamic conditions, target detections were not improved, and in the dynamic-slow task conditions, the average reaction time was greater in the low than in the high prevalence task condition, perhaps evidence of a necessity for a certain level of cognitive workload to be maintained in under-stimulating and uneventful conditions when search termination is not an option.

Interactions

Of particular interest for this project were potential interactions between target prevalence and the other conditions and their effects on target detection accuracy. As target prevalence is a known contributor to real-world misses in static searches, such as in radiology scans, understanding how prevalence interacts with time pressure and different search types in the lab would allow us to better predict and mitigate challenges in other, less studied search tasks, such as sonar and security monitoring. The results of our investigation, however, revealed no significant interactions between prevalence and display type or display time on target detection accuracy, as measured by hit rates, false alarms, sensitivity, or bias.

NASA-TLX

Rating of task demands on the NASA-TLX revealed that participants were sensitive to some manipulations' effects on cognitive demands and performance, but not others. The task performance effects of having less time to process and react to the display are evidenced in the NASA-TLX results, as observers rated the faster conditions as more temporally-demanding (Cohen's $d = .23$) and frustrating (Cohen's $d = .1$), as well as rating their performance as worse (Cohen's $d = .1$) when compared to slow conditions, aligning with their superior objective task performance in these task conditions. Despite the increased rating of temporal demands in the fast display conditions, participants did not rate faster conditions as more physically demanding, which perhaps indicates that misses were attributed less to physical or mechanical performance issues (e.g., clicking the mouse), and more to search behavior. These display time results overall serve as a manipulation check in which faster conditions were expected to yield greater temporal demand. However, when rating the effects of prevalence, participants thought that the low

prevalence condition was more frustrating than the high prevalence condition but did not perceive a difference in performance between the conditions. This last point is worth highlighting for its real-world importance and future research. While participants recognized performance decrements and demands due to some conditions, they did not perceive their performance to be lower in the low prevalence condition despite significantly lower objective performance. This worrying pattern suggests that developing a performance-based feedback system to make observers aware of their lower performance in low prevalence conditions may be necessary to mitigate the LPE by giving them the information they need to motivate behavioral adjustments.

Conclusion

The low prevalence effect, not commonly investigated in continuous dynamic display types, was robust and detrimental on search ability, reducing target detections by about 13%, averaged across conditions. These results provide evidence for Multiple Decision Model's decision criteria shifts affecting low target prevalence searches, even when quitting thresholds are not a factor, and supplement understandings of observer behavior and search performance. This finding also shows the LPE is equally concerning in dynamic and static searches, with and without time pressure, thus developing a mitigation strategy to reduce the LPE is warranted in all conditions we have explored. The NASA-TLX ratings indicate that while observers are unaware of the negative impact of certain conditions (e.g., low target prevalence) on their search performance, they are aware of the detrimental effects from other conditions (e.g., fast display speed); these findings provide valuable insight into observers' perceptions of their own performance and highlight the need for feedback mechanisms that make observers aware of hits and misses in real-time, allowing for self-correction to mitigate LPEs on target identification accuracy. An unexpected finding in which performance in dynamic display type tasks was not detected as significantly different from performance in static display type tasks (likely due to the predictability of the dynamic display's item motion) implies the use of different strategies in each search task. These conclusions assist the development of future studies; however, the conclusions are inferences concerning the search behavior exhibited within this study and require further investigation.

Limitations

Perhaps as a result of the current study being administered online through Amazon MTurk, a lack of control increased variability in the testing environment. Researchers, for example, were unable to ensure the same equipment was used, eliminate environmental distractions, or confirm that participants were not sharing attention between multiple tasks while completing the experiment. This online method of testing, however, allowed for greater accessibility, and greater diversity in participants and environments, which may increase generalizability to real-world visual search that lacks environmental and observer homogeneity.

The lack of *random* dynamic motion of the stimuli limits the real-world applicability of this work. Though our manipulation of motion with varying levels of prevalence and time pressure is a necessary step toward externally valid visual search research, the effects of predictable motion may not accurately reflect the challenges of a random motion task. In a task such as sonar, where stimuli are moving at different speeds and in random directions, a "sit and wait" search strategy is not possible, and observers would potentially need to track multiple objects, which is a known challenge for human observers (Pylyshyn & Storm, 1988). This

limitation prevents us from determining if prevalence and random motion interact to impact search performance.

Future Work

Future work could implement eye-tracking to enhance the comprehension of the current study's findings and verify our hypotheses concerning search behaviors and strategies. Assessing a third comparison group featuring a dynamic display with random item motion would be valuable for increasing generalizability of the study's findings and more closely reflecting the random item movement present in real-world visual search tasks (e.g., sonar monitoring). To understand search strategies applied in an actual sonar environment, measuring performance of professional technicians monitoring sonar displays should be considered. Furthermore, increasing the length of the visual search task to more closely resemble the length of a real-world watch times such as sonar monitoring could yield new insights into common detriments to visual search performance and target identification accuracy attributed to time-on-task fatigue (Peltier, Daley, Guillory, Handy, & Wilson, 2022).

By investigating the effects of various search conditions on visual search performance that may affect sonar, we have found that the LPE and observers' perceived performance raise concerns meriting further investigation. LPE-remediating strategies are not currently used in real-world dynamic searches such as sonar or security monitoring, but given that the present work demonstrates that the LPE is of similar magnitude—and possibly even greater magnitude without the potential confounding predictability effects on search performance—in both static and dynamic search tasks, investigation of real-world strategies to improve observer performance overall is necessary.

References

- Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, 14, 427-432.
- Allen, R., McGeorge, P., Pearson, D., & Milne, A. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, 18, 337–347. doi:10.1002/acp.975
- Allen, R., McGeorge, P., Pearson, D., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology*, 59, 1101–1116. doi:10.1080/02724980543000097
- Alvarez, G., Konkle, T., & Oliva, A. (2007). Searching in dynamic displays: Effects of configural predictability and spatiotemporal continuity. *Journal of vision*, 7(14), 12-12.
- Beck, M., Peterson, M., & Vomela, M. (2006). Memory for where, but not what, is used during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 235.
- Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods*, 50(3), 937-962.
- Boot, W. R., Becic, E., & Kramer, A. F. (2007). Temporal limitations in multiple target detection in a dynamic monitoring task. *Human Factors*, 49(5), 897-906.
- Boot, W. R., Kramer, A. F., Becic, E., Wiegmann, D. A., & Kubose, T. (2006). Detecting transient changes in dynamic displays: The more you look, the less you see. *Human Factors*, 48(4), 759-773.
- Burg, A. (1966). Visual acuity as measured by dynamic and static tests: A comparative evaluation. *Journal of Applied Psychology*, 50(6), 460–466. doi:https://doi.org/10.1037/h0023982
- Erikson, R. (1964). Visual search performance in a moving structured field. *Journal of the Optical Society of America*, 54, 399–405.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods*, 175–191.
- Geyer, T., von Mühlenen, A., & Müller, H. J. (2007). What do eye movements reveal about the role of memory in visual search? *Q. J. Exp. Psychol.*, 60, 924–935.
- Godwin, H., Menneer, T., Riggs, C., Cave, K., & Donnelly, N. (2015). Perceptual failures in the selection and identification of low prevalence targets in relative prevalence visual search. *Attention, Perception, & Psychophysics*, 77, 150 –159.

- Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.
- Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of educational statistics*, 17(4), 315-339.
- Hautus, M. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behav Res Methods Instrum Comput*, 46–51.
- Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, 39(1), 30-47.
- Hoang, I., Ranchet, M., Derollepot, R., Moreau, F., & Paire-Ficout, L. (2020). Measuring the cognitive workload during dual-task walking in young adults: A combination of neurophysiological and subjective measures. *Frontiers in Human Neuroscience*, 14, 511.
- Hofer, F., & Schwaninger, A. (2005). Using threat image projection data for assessing individual screener performance. *WIT Transactions on the Built Environment*, 82, 417– 426.
- Horowitz, T. (2017). Prevalence in visual search: From the clinic to the lab and back again. *Japanese Psychological Research*, 59(2), 65-108.
- Horowitz, T., & Wolfe, J. (1998). Visual search has no memory. *Nature*, 394(6693), 575-577.
- Hout, M., Walenchok, S., Goldinger, S., & Wolfe, J. (2015). Failure of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 977–994. doi:10.1037/xhp0000053
- Hulleman, J. (2009). No need for inhibitory tagging of locations in visual search. *Psychonomic Bulletin & Review*, 16-120.
- Ishibashi, K., Kita, S., & Wolfe, J. (2012). The effects of local prevalence and explicit expectations on search termination times. *Attention, Perception, & Psychophysics*, 72, 115–123.
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, H. Hotelling, & e. a. (eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278-292). Stanford University Press.
- Makowski, D. (2018). The psycho package: an efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 470.

- Mansikka, H., Virtanen, K., & Harris, D. (2019). Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*, 62(2), 246-254.
- Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect: Visual search for exceedingly rare items is highly susceptible to error. *Psychological science*, 284-289.
- Nelson, T. D., James, T. D., Nelson, J. M., Johnson, A. B., Mason, W. A., Yaroch, A. L., & Espy, K. A. (2020). Associations between specific components of executive control and eating behaviors in adolescence: A study using objective and subjective measures. *Appetite*, 154(1), 1-10.
- Peltier, C., & Becker, M. (2015). Memory is necessary in visual search with limited guidance. *Journal of Vision*, 15(12), 1357. doi:10.1167/15.12.1357
- Peltier, C., & Becker, M. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1466.
- Peltier, C., & Becker, M. W. (2017). Target-present guessing as a function of target prevalence and accumulated information in visual search. *Atten Percept Psychophys*, 79, 1064–1069.
- Peltier, C., Daley, M., Guillory, S., Handy, J. D., & Wilson, B. (2022). Vigilance Performance Declines Faster When Monitoring for a Signal in Two Modalities Compared to One. *Advances in Cognitive Psychology*.
- Posner, M. I., Rafel, R. D., Choate, L. S., & Vaughn, J. (1985). Inhibition of return: Neural basis and function. *Cognitive neuropsychology*, 211-228.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3), 179-197.
- Ratcliff, R., & Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356. doi:http://dx.doi.org/10.1111/ 1467-9280.00067
- Rich, A., Kunar, M., Van Wert, M., Hidalgo-Sotelo, B., Horowitz, T., & Wolfe, J. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, 8, 1-17. doi:http://dx.doi.org/10.1167/8.15.15
- Rieger, T., Heilmann, L., & Manzey, D. (2021). Visual search behavior and performance in luggage screening: Effects of time pressure, automation aid, and target expectancy. *Cognitive Research: Principles and Implications*, 6(1), 1-12.
- Rizzi, A., Buck, J., & Anderson, V. (1979). Performance effects of variables in dynamic visual inspection. *AIEE Transactions*, 11(4), 278-285.
- Schwark, J., Sandry, J., Macdonald, J., & Dologov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention, Perception, & Psychophysics*, 74, 1583–1589. doi:http://dx.doi.org/10.3758/s13414-012-0354-4

- Software, M. (2010). Retrieved from <https://www.millisecond.com/>
- Stothart, C., Clement, A., & Brockmole, J. (2018). Satisfaction in motion: Subsequent search misses are more likely in moving search displays. *Psychonomic bulletin & review*, 25(1), 409-415.
- Tombu, M., & Seiffert, A. (2008). Attentional costs in multiple-object tracking. *Cognition*, 108, 1-25. doi:doi:10.1016/j.cognition.2007.12.014
- Van Wert, M., Horowitz, T., & Wolfe, J. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics*, 71(3), 541-553.
- von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, 14(4), 309-314.
- Wang, L., & Sun, X. (2015). Prevalence Effects in X-Ray Screening Tasks with a Static or Dynamic Visual Display: Is There Any Difference? *International Conference on Engineering Psychology and Cognitive Ergonomics*, 77-83.
- Wang, Z., Zhang, K., & Klein, R. (2010). Inhibition of return in static but not necessarily in dynamic search. *Attention, Perception, & Psychophysics*, 72(1), 76-85.
- Wentworth, R., & Buck, J. (1982). Presentation effects and eye-motion behaviors in dynamic visual inspection. *Human Factors*, 24(6), 643-658.
- Wickens, T. (2002). *Elementary signal detection theory*. New York: Oxford UP.
- Williams, L., & Borow, M. (1963). The effect of rate and direction of display movement upon visual search. *Human Factors*, 5(2), 139-146.
- Wolfe, J., & Van Wert, M. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, 20, 121-124. doi:http://dx.doi.org/10.1016/j.cub.2009.11.066
- Wolfe, J., Horowitz, T., & Kenner, N. (2005). Rare items often missed in visual searches. *Nature*, 435(7041), 439-440.
- Wolfe, J., Horowitz, T., Van Wert, M., Kenner, N., Place, S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, 136(4), 623.
- Yang, L., Song, F., Zhang, Y., & Yu, R. (2016). Effects of target prevalence, target presence and time pressure on visual search performance. *Industrial Engineering Journal*, 19(6), 83.
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive psychology*, 295-340.
- Yantis, S., & Gibson, B. (1994). Object continuity in apparent motion and attention. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 48(2), 182.

- Yu, R., & Chan, A. (2015). Display movement velocity and dynamic visual search performance. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(3), 269-278.
- Yu, R., Yang, L., Guo, X., & Zhang, Y. (2015). Effect of time pressure on dynamic visual search performance. *Effect of time pressure on dynamic visual search performance. Procedia Manufacturing*, 3, 4658-4664.

Appendix A: NASA Task Load Index Rating Scale Definitions

RATING SCALE DEFINITIONS			
TITLE	ENDPOINTS	SURVEY QUESTION	DEFINITIONS
MENTAL DEMAND	1-10 <i>Low/High</i>	How mentally demanding was the task?	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	1-10 <i>Low/High</i>	How physically demanding was the task?	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	1-10 <i>Low/High</i>	How hurried or rushed was the pace of the task?	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	1-10 <i>Low/High</i>	How successful were you in accomplishing what you were asked to do?	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	1-10 <i>Perfect/Failure</i>	How hard did you have to work to accomplish your level of performance?	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	1-10 <i>Low/High</i>	How insecure, discouraged, irritated, stressed, and annoyed were you?	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?