

The Viability of Web-based Eye Tracking

by

Amy vanWell

B.Sc., University of Victoria, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Amy vanWell, 2023
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

We acknowledge and respect the lək'wəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ people whose historical relationships with the land continue to this day.

The Viability of Web-based Eye Tracking

by

Amy vanWell

B.Sc., University of Victoria, 2021

Supervisory Committee

Dr. James Tanaka, Supervisor
Department of Psychology

Dr. Jordana Wynn, Departmental Member
Department of Psychology

Abstract

The recent COVID-19 pandemic has forced vision scientists to develop innovative, online solutions to their research questions. After the introduction of new open-source software and improvements in the personal laptop hardware, researchers have been able to recreate increasingly sophisticated measures online historically only measured in-person. In this paper, we explore the boundaries of online research to describe the development and testing of a new web-based eye tracking system, “*Gazer*”. *Gazer* is an accessible eye tracking system for vision scientists that harness the open-source *Webgazer* to record screen-based gaze locations, using the cameras present in participant’s personal laptop. We directly compare *Gazer* to the established Eyelink 1000 in two separate experiments, and determine it has comparable temporal and spatial precision for recording fast exogenous eye movements, and endogenous movements. In a final experiment, we record the development of gaze strategies in a repeated remote visual search task to indicate *Gazer*’s potential to address sophisticated attentional research questions. Overall, we present *Gazer* as a viable method for researchers to conduct gaze-based research online.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vii
Acknowledgements	ix
Chapter 1: General Introduction	1
Chapter 2: Gazer: A New Tool for Web-based Eye Tracking	6
Chapter 3: Where's Waldo? Exploring Gaze Strategy in a Visual Search Task Online and In-Person	36
Chapter 4: Evidence for Visual Search Performance Improvement through Perceptual Training	58
Chapter 5: General Discussion	78

List of Tables

Chapter 2

Table 1. Absolute Coordinates of the Eight Red Dot Targets as Compared to the Calculated Clusters of Gaze Recorded by both Gazer and the Eyelink 1000. Clusters were Calculated as the Average Gaze Locations in the Final 1000ms of Trials. Coordinates Locations are Reported using Screen Proportion as a Metric.....26

Chapter 3

Table 1. Behavioural (Reaction Time and Accuracy) Measures Recorded During a Web-based Where's Waldo Experiment.....43

Table 2. Behavioural (Reaction Time and Accuracy) Measure for an In-lab Where's Waldo Experiment.....48

Chapter 4

Table 1. Slope and Intercept for Reaction Time to Press the [space] key, Conditioned on Session.....65

Table 2. Fixations Per Trial, Fully Conditioned (*i.e.*, Intercept and Slope) on Session.....67

Table 3. Rate (Fixations Per Second) Partially Conditioned (Slope Only) on Session.....68

List of Figures

Chapter 1

Figure 1. An Eyelink 1000 Device, a Common In-lab Eye Tracking System that Uses the Corneal Reflection Method and a Close-Up Image of an Eye with a Present Corneal Reflection.....	2
Figure 2. A 3D Face Mesh Shown by Itself (left) and Superimposed onto a Gazer Video Feed (right).....	3

Chapter 2

Figure 1. An Example of a 3D Face Mesh Before Superimposition onto a Gazer Video Feed...	10
Figure 2. A Sample Image of the Bounding Box Provided to Participants During Calibration...	14
Figure 3. A Sample Sequence of Obtaining Calibration Data After a Click on a Calibration Stimulus, a Component of the “Point-and-Click” Calibration. Participant Clicks on a Stimulus (left) and Immediately After the Location and Face Orientation from the Time of the Mouse Click are Saved (right).	15
Figure 4. The Timeline of One Trial in the Attentional Cueing Task.....	17
Figure 5. The Distance to Target Recordings for the Gazer Predictions of Individual Participants (left) and Aggregated Across All Participants (right).....	19
Figure 6. Average Gaze Location Predictions as Recorded by Gazer in the First 1000 ms (left) and Final 1000 ms (right) for Trials with Each Target Location. Average Gaze Locations separated by Target Position (bottom).....	20
Figure 7. The Distance to Target Recordings for the Eyelink for Individual Participants (left) and Aggregated Across All Participants (right).	24
Figure 8. Average Gaze Location Predictions as Recorded by Eyelink in the First 1000 ms (left) and Second 1000 ms (right) for Trials by Target Location. Average Gaze Locations by Location in the Final 1000 ms (bottom).	25

Figure 9. Distance to Target Measures for All Participants Recorded by Eyelink (left) and Gazer (right).....	28
---	----

Chapter 3

Figure 1. A Sample Sequence of a Where's Waldo Trial.....	41
Figure 2. Two Example Backgrounds Implemented in the Where's Waldo Task.....	44
Figure 3. Euclidean Distance Between Participants Average Gaze Location Recorded by Gazer in 20 ms Time Bins and the Target Waldo in the 5000 ms Immediately Before the [space] Key Press was Made.....	46
Figure 4. Euclidean Distance Between Participants Average Gaze Location Recorded by Eyelink in 20 ms Time Bins and the Target Waldo in the 5000 ms Immediately Before the [space] Key Press was Made.....	50

Chapter 4

Figure 1. An Example of a Background Used in a Where's Waldo trial, with a Waldo Target Superimposed on the Top Left of the Background.....	61
Figure 2. An Example of a Trial Timeline for the Where's Waldo Task.....	63
Figure 3. Average Reaction Time Per Trial Performed at Sessions 0,1, and 2.....	69
Figure 4. Average Number of Fixations Per Trial Performed at Sessions 0,1, and	70
Figure 5. Average Number of Fixations Made Per Trial in Sessions 0 (left), 1 (middle), and 2 (right).....	73

Acknowledgements

I'd like to express my deep gratitude to my supervisor Jim for his support during both my graduate and undergraduate years at UVic. I'm very privileged to have been the recipient of your time and effort, and I'll always be grateful for both your friendship and the work I've been able to conduct with your guidance. I'd also like to extend my appreciation to Dr. Jordana Wynn for her willingness to give up precious time and serve on my committee, and of course my thanks to all the friends and research assistants I've had the pleasure to work with as a member of the Different Minds Lab.

To my dearest Mom and Dad – I love you both to the ends of the Earth. Mom, thank you for your endless love and kindness. Dad, you always have been and always will be my Hero.

To my sis and JI – come home!

To BB – love and appreciate you forever.

Chapter 1: General Introduction

After a sudden and immediate requirement for online data collection, it is no surprise that a need for more sophisticated web-based data has also emerged. However, while it is well-established that online experiments yield behavioural data (*e.g.*, accuracy, reaction time) that is comparable to in-lab data (Anwyl-Irvine *et al.*, 2020; de Leeuw, 2015), the fidelity of more complicated measures is not known. The work presented in this thesis describes the “*Gazer*” project conducted at the University of Victoria to develop and test an eye-tracking system that operates remotely through participant’s browsers. We aim to expand the accessibility of eye movement data to vision scientists with the establishment of an open-source, straightforward, and online eye tracking system.

In this thesis, I assess the viability of the *Gazer* eye tracking system by directly comparing its precision for two different components of visual attention to the industry standard in-person Eyelink 1000. Critically, the direct comparison evaluates differences in the quality of data produced using the indirect eye tracking method employed by *Gazer*, and the direct corneal reflection method used by Eyelink.

Direct versus Indirect Eye Tracking

The most common eye tracking systems employ a corneal reflection method in which an infrared light is directed towards the center of the eyes (pupil), creating reflections in both the pupil and the cornea. The position of the corneal reflection, relative to the location of the pupil, is continually measured during experimentation to precisely calculate gaze location at a frequency of 1000Hz (maximum) and a spatial resolution of 0.1 degree (or lower).

Figure 1

An Eyelink 1000 Device, a Common In-lab Eye Tracking System that Uses the Corneal Reflection Method (left) and a Close-Up Image of an Eye with a Present Corneal Reflection (right).



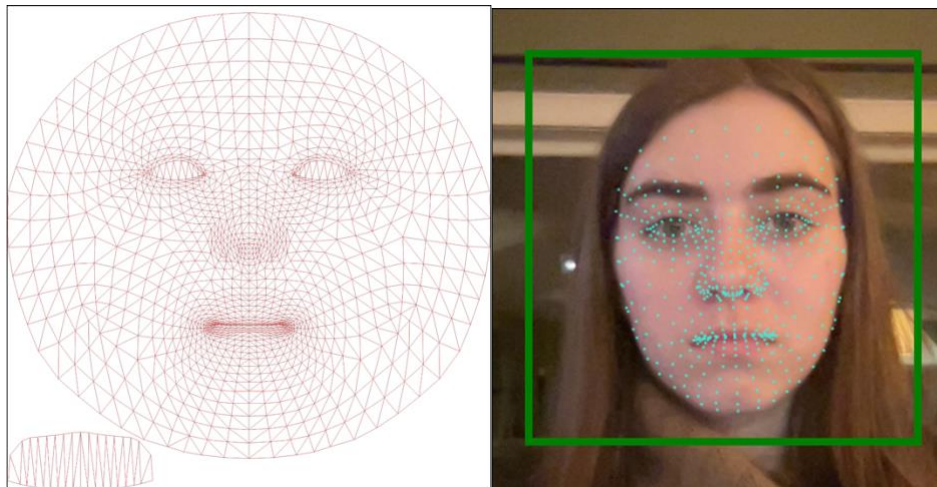
Note. Infrared light is generated and beamed from the right side of the device while a video recording is made using the camera on the left. The location of the reflection (indicated by an arrow), relative to the pupil, is used to calculate the eye gaze location of participants.

As an alternative to in-lab systems, we have developed the web-based eye tracking *Gazer* system that incorporates the open-source *Webgazer* software (Papoutsaki, 2015) into the jsPsych experimental framework. *Gazer* accesses the raw video feed from participant's laptops and overlays a 3D face mesh (Figure 2, left) onto each participant's face. The individual points of the mesh map onto specific facial features of the face, and the distortions of their locations produced during the mapping provide information on the face's orientation. During a "Point-and-Click" self-calibration, participants train a Bayesian algorithm to associate orientations of the face mesh with specific gaze locations on the screen. During the experiment, gaze location is continually

predicted at a frequency of 30Hz (*i.e.*, one time sample every 33 ms). The *Gazer* system was developed to expand existing software for online eye-tracking into a format that is customized for vision science and accessible without extensive programming experience.

Figure 2

A 3D Face Mesh Shown by Itself (Left) and Superimposed onto a Gazer Video Feed (Right).



Summary

Corneal reflection systems and the Gazer program take two different approaches to eye tracking recording. The corneal reflection method is “direct” where gaze location is based on actual measurements of the eye including the relative distance between the pupil location and corneal reflection. In contrast, the Gazer system is an indirect measure that uses whole-head information to estimate gaze location based on a Bayesian analysis.

Eye Movements and Attention

It is common for attention researchers to assume participants “pay attention” to the screen locations they fixate on. Therefore, gaze locations recorded using remote and in-person eye tracking methods are used to operationalize visual attention. In this thesis, we focus on two components of visual attention that generate eye movements: “exogenous” and “endogenous” attention. Exogenous attention is stimulus-driven by a salient characteristic (*e.g.*, a bright light) and generates rapid eye movements to the stimulus (Posner, 1980). Endogenous attention is driven by top-down processing (*i.e.*, the current task or goal) and results in eye movements patterns that are less directly towards to the target (Jonides, 1981). Eye tracking methods also calculate fixations as sustained periods of gaze on a single location. Critically, the distribution of fixations over a single viewing of a stimulus can inform what is salient about an image (*i.e.*, visually “stands out” to the viewer) and therefore describe what features capture visual attention. In this thesis, we evaluate *Gazer*’s precision to record exogenous eye movements, endogenous eye movements, and fixations.

Summary of Research

The goal of the first two experiments was to quantify the precision of *Gazer* for recording rapid, direct eye movements (*i.e.*, saccades) driven by exogenous attention (Chapter 2) and the less straight-line eye movements driven by top-down endogenous attention (Chapter 3) by directly comparing its results to the EyeLink 1000 in an identical task. After establishing *Gazer*’s precision to record eye movements we use it to investigate the development of gaze strategies within a repeated visual search task (Chapter 4).

References

- Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo K. Evershed. 2020. "Gorilla in Our Midst: An Online Behavioral Experiment Builder." *Behavior Research Methods* 52 (1): 388–407.
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movements. *Attention and Performance*, 0. <https://cir.nii.ac.jp/crid/1572261550630775936>
- Leeuw, Joshua R. de. 2015. "JsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser." *Behavior Research Methods* 47 (1): 1–12.
- Papoutsaki, Alexandra. 2015. "Scalable Webcam Eye Tracking by Learning from User Interactions." In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 219–22. CHI EA '15. New York, NY, USA: Association for Computing Machinery.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25.

Chapter 2: Gazer: A New Tool for Web-based Eye Tracking

Amy vanWell¹ & James W. Tanaka¹

1. Department of Psychology, University of Victoria

Abstract

The evolution of personal laptop cameras and eye tracking software have created an opportunity to explore new possibilities in web-based science. In this paper, we describe ‘*Gazer*’ – a web-based program that allows researchers the capability to record eye-tracking behaviors using their internet browser and the laptop’s camera. In the current study, we compared the temporal and spatial performance of the web-based *Gazer* program to the laboratory-based Eyelink 1000 in an attention task using an exogenous cue (Posner, 1980). On each trial, a gaze-contingent cross was presented at the centre of the screen. After participants maintained their gaze on the cross for 500 *ms*, a target red dot appeared in one of eight peripheral locations. Participants were instructed to redirect their gaze to the target red dot and maintain their gaze position at the target location until the target disappeared, approximately 2 seconds later. Our results indicated that on both systems, participants reliably shifted their gaze towards the target within 400 *ms* of target onset. An analysis of recorded clusters of gaze coordinates corresponded to target locations and the positions did not reliably differ across the two systems. Our findings show that eye tracking data recorded by *Gazer* was comparable in its temporal and spatial precision, however, eye movements recorded by *Gazer* were lagged behind and more variable than Eyelink. We propose *Gazer* system as an alternative to high-end, laboratory-based systems; thus, opening new avenues of research for behavioural scientists.

Introduction

Over the last decade, a growing number of experimental psychologists have turned to online experimental platforms (e.g., jsPsych, PsychoPy, OpenSesame) to conduct behavioural research. For researchers, online methods provide ready access to a large and diverse participation pool and afford rapid data collection without the need of a physical lab or a team of research assistants. Critically, web-based experiments have been shown to yield response time and accuracy data that is comparable to in-lab experiments in terms of their precision and reliability (Anwyl-Irvine *et al.*, 2021; de Leeuw, 2015).

Here, we broaden the experimental capabilities of web-based research by introducing an innovative, accessible eye tracking system called *Gazer*. Built on the existing *Webgazer* program (Papoutsaki, 2015), *Gazer* is an open-source system in eye tracking that is implemented in the popular jsPsych experimental framework (de Leeuw, 2015). In the described experiments, we compare the spatial and temporal performance of *Gazer* to the performance of the lab-based, research-grade Eyelink 1000 system. Our findings indicate that the web-based *Gazer* is comparable to the lab-based system in terms of its spatial precision and temporal precision and therefore, we conclude that it is a viable and available research alternative to lab-based systems. In the current study, we quantify spatial precision as the screen distance between gaze coordinates as recorded by Eyelink and *Gazer* and the precise location of the target stimulus on the screen. Temporal precision is defined as each system's capability to record the onset of an eye movement towards the target after its sudden appearance.

Direct Corneal Reflection versus Indirect “Whole Head” Approaches to Eye Tracking

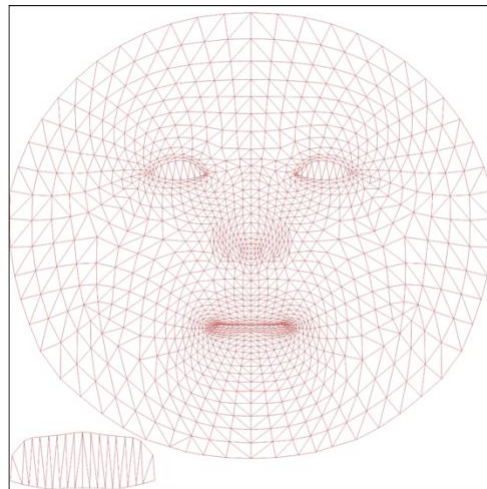
The research standard in eye tracking systems employs a corneal reflection method in which an infrared light is directed towards the center of the eyes (pupil), causing detectable reflections in both the pupil and the cornea (the outermost optical element of the eye). During the calibration process, the distance between the corneal reflection and pupil are mapped to observer’s points of fixation on the screen. The position of the corneal reflection, relative to the location of the pupil, is then continually measured and updated during experimentation to precisely calculate gaze location. Gaze locations are monitored at a frequency of 1000Hz (maximum) and a spatial precision of 0.1 degree (or lower), yielding data at a high temporal and spatial precision. Given the high spatial precision and temporal precision of corneal reflection systems, they have become the research standard for conducting eye movement experiments in such domains as reading (Jarodzka & Brand-Gruwel, 2017), face recognition (Alonso *et al.*, 2017; Wu *et al.*, 2012), and decision-making (Fiedler & Glöckner, 2012; Gidlöf *et al.*, 2013; Jenke *et al.*, 2021). However, despite their superior measurement capabilities, corneal reflection systems have certain practical limitations for the researcher. Corneal reflection systems are expensive (*e.g.*, averaging around \$17,000 USD), and require the operation of a trained research assistant. They are also a physically bulky piece of equipment that requires a separate, dedicated room for testing. Furthermore, the calibration process can be temperamental and time-consuming if the participant is improperly positioned, displays excessive head movements, or fails to attend to fixation points.

As an alternative, we developed the web-based *Gazer* eye-tracking program that avoids some of the drawbacks and inconveniences associated with lab-based systems. Built on the existing *Webgazer* program (Papoutsaki, 2015), *Gazer* utilizes “whole head” information, in

contrast to the direct corneal reflection (*i.e.*, direct to eye) method. During *Gazer* experimentation, a 3D face mesh (as shown in Figure 1) is superimposed onto participant's faces based on webcam input to estimate the observer's head orientation and gaze direction. During self-calibration, participants train a Bayesian algorithm to associate orientations of the face mesh with specific gaze locations on the screen. Hence, in contrast to retinal reflection systems, gaze location is not measured directly from the observer's eyes and is inferred from their general head position. During online experimentation, gaze location is continually predicted at a frequency of 30Hz (*i.e.*, one time sample every 33 ms).

Figure 1

An Example of a 3D Face Mesh Before Superimposition onto a Gazer Video Feed.



Note. The *Gazer* program maps the individual points of the face mesh onto features of the face in the video feed, and the distortion of their locations after being placed onto the video feed provide input on the orientation of participant's facial features. Source:

https://docs.snap.com/assets/images/face-mesh_face-mesh-a11-5ad1f789e6dc8a6b15a2b54ac7d1ba81.png

Semmelman and Weigelt (2018) tested the efficacy of the *Webgazer* system across three experimental paradigms: 1) a fixation task in which participants fixated on a target circle, 2) a pursuit task in which participants tracked the movement of a target stimulus, and 3) a face task in which participants freely viewed a face for 3000 *ms* to record regions of interest. The main finding was that on the three tasks, the *Webgazer* program yielded data that were comparable to eye tracking data reported in the literature using more sophisticated corneal reflection systems (Semmelmann & Weigelt, 2018). Although the results of Semmelmann and Weigelt study are promising, they noted that there were several areas where their web-based eye tracking system could be improved. First, they found that the calibration process could be long, taking up to 50% of the experimental time of 30 minutes. Second, the difficulty of calibration might explain the relatively high incompleteness rate in their study where only 32 of the 84 participants who began the study successfully submitted a full data set. Finally, to validate the temporal and spatial precision of the web-based eye tracking system, it would be desirable to test its output to the output produced by a lab-based, corneal reflection system using the same experimental task (Semmelmann & Weigelt, 2018).

The goals of the current study were two-fold: First, we were interested in comparing the performance of the *Gazer* system to lab-based systems with respect to its temporal and spatial precision. Secondly, we wanted to evaluate the robustness of the *Gazer* system when implemented across a large sample of users using different devices under different testing conditions.

Experiment 1A: Exogenous Attention Task Using the Gazer System

It is an axiom in cognitive psychology that “where the eye is looking is where the mind is attending”. In attentional research, eye movements have been shown to be a reliable indicator of spatial attention where stimulus information located at the point of fixation is processed faster and more accurately ((Bashinski & Bacharach, 1980; Downing, 1988); (Hawkins *et al.*, 1990). Spatial attention has been broadly divided into two types: exogenous attentional mechanisms and endogenous attentional mechanisms. Exogenous attention is stimulus-driven such that our attention is automatically captured by a salient visual feature or stimulus event. For example, the sudden appearance of bright white light generates an exogenous eye movement towards the abrupt onset of the stimulus event. In a speeded saccade task with no distractors, exogenous eye movements can be made in less than 240 *ms* (van Zoest *et al.*, 2004).

In Experiment 1A, we investigated the capability of the *Gazer* program to track eye movement behaviours in an exogenous cueing task. The aim of the experiment was to measure the spatial and temporal correspondence between the onset of the eye movement towards the target and the onset of the target stimulus, and evaluate its robustness across participants, their testing conditions, and devices.

Method

Participants

Forty-one participants (number of females = 37) ranging in age from 19 to 43 ($m = 23.9$ years) took part in this study in exchange for course credit. A total of 42 participants attempted calibration, and 41 successfully reached the threshold and proceeded to the trials. All participants

had normal or corrected-to-normal vision with contacts. The experiment was approved by the Research Ethics of the University of Victoria.

Materials and Equipment

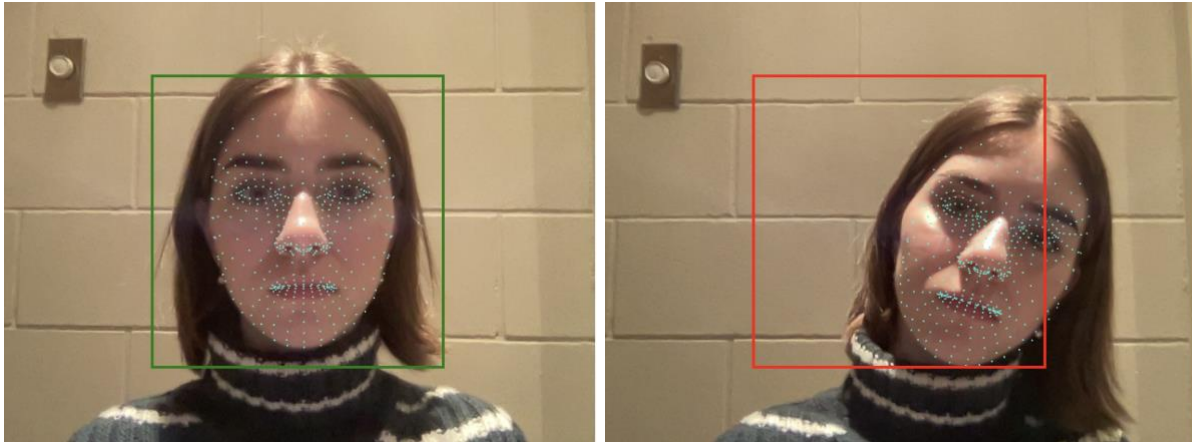
The experiment was conducted remotely in the browser. Participant's hardware was limited to standard personal laptops with built-in video cameras, and software was limited to the Chrome browser. The experiment was hosted on Heroku and accessed through a URL, and all data was saved to a MongoDB database hosted on a Digital Ocean server.

Procedure

To properly set up the computer to run the *Gazer* program, participants were instructed to position their laptop so that the webcam was located at eye height. To maintain a consistent viewing distance of approximately 34 centimeters, participants were asked to place their face within the bounding box such that their forehead was aligned with the top of the box and their chin with the bottom of the box (see Figure 2). Proper face placement was indicated by a green border around the box and incorrect face placement was indicated by a red border around the box.

Figure 2

A Sample Image of the Bounding Box Provided to Participants During Calibration.



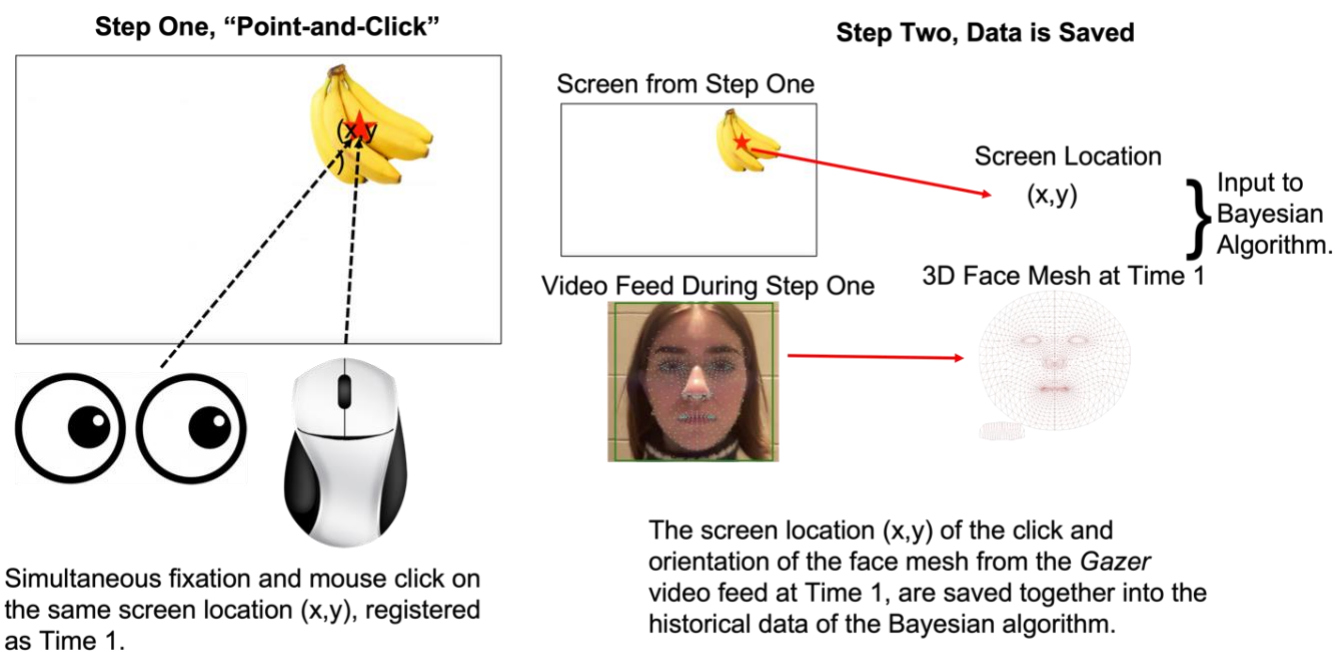
Note. The bounding box is bordered in green (left) when a face is detected, and red when no face is correctly present (right).

After their face was properly aligned with the camera, participants began the self-calibration phase of the study. *Gazer* employs a “*point-and-click*” method for calibration, during which participants are sequentially presented a series of calibration stimuli (bunches of bananas, approx. 285 px by 325 px) at eight peripheral locations in clockwise order beginning at the top, lefthand corner of the screen. Participants were instructed to maintain their fixation on the center red star of each calibration stimulus while clicking on the red star with their mouse five times. Participants are instructed to use their mouse to click on each banana stimulus 5 times (after which it will disappear) to calibrate the eye-tracker to those locations. Participants click on a total of four banana stimuli, appearing in clockwise order and starting in the upper left-hand corner. After clicking on all the banana images, they are asked to click on a central cross another 5 times before proceeding to validation.

With each mouse click, the pixel location and the orientation of the face mesh was saved together, as shown in Figure 3. Critically, the pairs of face mesh orientations and pixel locations collected at each click provide the historical data for real-time gaze prediction during the experiment (Figure 3, right). After following the point-and-click procedure for four stimuli, participants proceeded to validation.

Figure 3

A Sample Sequence of Obtaining Calibration Data After a Click on a Calibration Stimulus, a Component of the “Point-and-Click” Calibration. Participant Clicks on a Stimulus (left) and Immediately After the Location and Face Orientation from the Time of the Mouse Click are Saved (right).



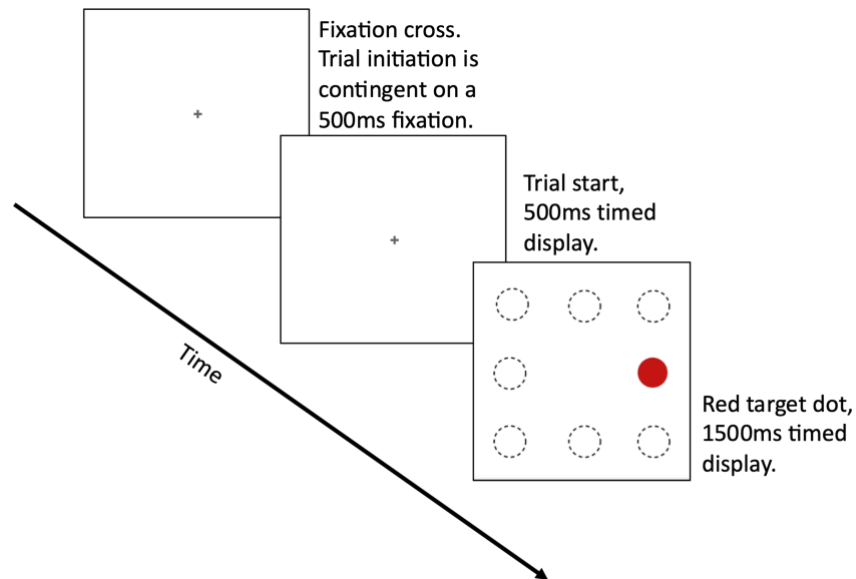
During the validation phase, participants fixated on a cross located at the center of the screen for five seconds. During this time, their predicted gaze locations were collected and

compared to the true screen coordinates. Accuracy of the calibration was calculated as the average distance between each prediction point and the center of the screen (*e.g.*, a 90% calibration accuracy recorded points to be on average within 10% distance to the center). If the difference between the predicted and actual locations was above a 70% accuracy threshold the participant continued to the experimental phase. If accuracy was below threshold the calibration process was repeated; the calibration process could be repeated a maximum of three times.

Experimental Task. At the beginning of each trial, participants were presented with a gaze-contingent fixation cross. Participant were asked to fixate on the cross for 500 *ms* plus an additional 500 *ms* (1000 *ms* in total) and once they had met this criterion, the cross was replaced by the target circle stimulus (50 *px* by 50 *px*, not scaled by screen resolution), and the attentional cueing task was initiated. The target stimulus was presented at one of eight possible locations around the perimeter of the central fixation (as shown in Figure 4). Participants were instructed to move their eyes to the target circle and fixated on the target circle until the target disappeared approximately 2 seconds later. The experimental trials were presented in five blocks of eight trials (40 trials in total) with the target appearing at one of eight randomly selected locations within a block.

Figure 4

The Timeline of One Trial in the Attentional Cueing Task.



Note. Participants performed gaze-contingent initiation on the fixation cross with a threshold of 500 *ms*. After initiation, they focused on the cross for an additional 500 *ms*, then were asked to shift their gaze to an exogenous cue (50 *px* by 50 *px* red dot) that appeared at one of eight locations.

Data Analysis

Gazer Data Preparation. From *Gazer*, the raw eye tracking output consists of an array of x,y coordinates (in pixel units) that are recorded at 33 *ms* intervals (30*Hz*). To prepare the data for analysis, the trial outputs were compiled into a longform dataframe in which the columns included target position (*i.e.*, the location of the red target dot labelled from 1-8), x-coordinate, y-coordinate, and timestamp. Due to the variability in screen resolution, pixel (*px*) coordinates were converted to be proportional to create a common metric across screens. For example, if a participant had a screen precision of 1000 *px* by 1080 *px* and a gaze coordinate was recorded at

the coordinates $x = 350 \text{ px}$ and $y = 500 \text{ px}$, the proportion values would be $(350/1000) \text{ px}$ by $(500/1080) \text{ px}$ or 0.35 px by 0.46 px . Using this method, comparisons can be made more readily across screen sizes. This proportion measure deviates from the typically reported visual angle, however, as a limitation of remote experimentation the distance between the participant and the screen is not known for certain, and therefore visual angle is not estimable.

Distance to Target Calculation. From the raw data of all participants, gaze coordinates (in pixels) were averaged within 20 ms time bins. The straight-line distance was calculated between the average coordinates in each time bin, and the location of the red target stimulus. Following the method of Semmelmann & Weigelt (2018), we used a straight-line (Euclidean) distance between gaze predictions and the target to demonstrate how gaze shifts relative to the target. For example, if a target was placed at the coordinates $(500 \text{ px}, 500 \text{ px})$, and gaze was recorded at $(300 \text{ px}, 350 \text{ px})$, the straight-line distance would be calculated as:

$$\sqrt{((500-300)^2 + (500-350)^2)} = 250 \text{ px distance.}$$

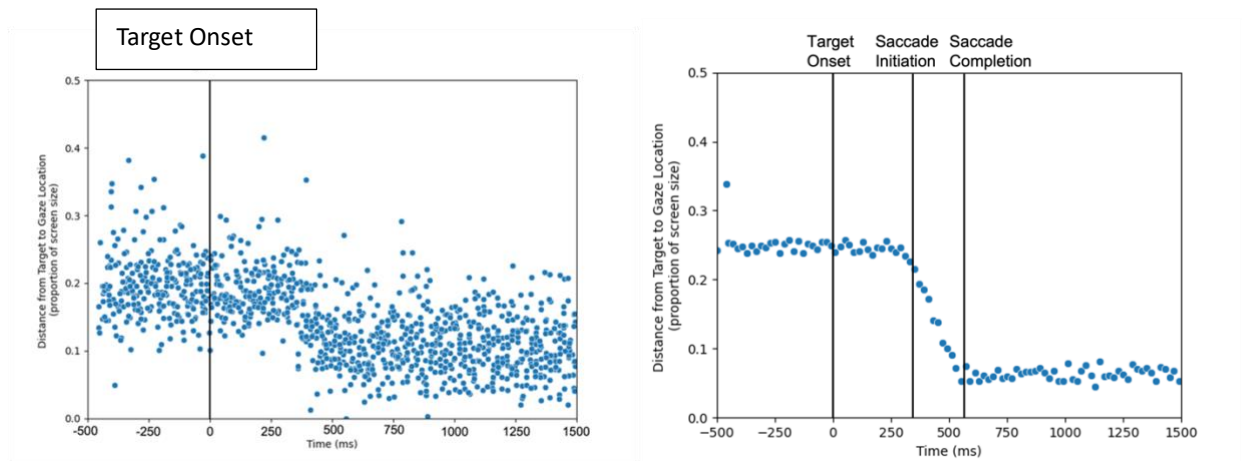
Results

Temporal Precision

The Euclidean distance between the red target dot and participant's average gaze locations is visualized in Figure 5. An abrupt shift in the spatial location of the participant's gaze was present after the onset of the target at 500 ms and was interpreted as a saccade. Saccade onset was calculated as the shift from zero a negative slope and occurred approximately 370 ms after the target onset.

Figure 5

The Distance to Target Recordings for the Gazer Predictions of Individual Participants (left) and Aggregated Across All Participants (right).



Note. Each dot represented the average straight-line distance between the target and participant's average gaze in a 20 ms time bin. Time is shown in 20 ms bins on the x-axis, and distance to target is shown on the y-axis. A horizontal, zero slope represents a sustained gaze location, and a non-zero slope indicates an eye movement towards (negative slope) or away from (positive slope) the target. The average speed of the saccade was 192 ms (562 ms eye movement completion – 370 ms onset).

The onset and completion of the eye movement towards the target occurred an average of 370 ms and 562 ms after target onset. Hence, the average time window of the saccade, calculated as the time between the onset and completion of the eye movement, was 192 ms.

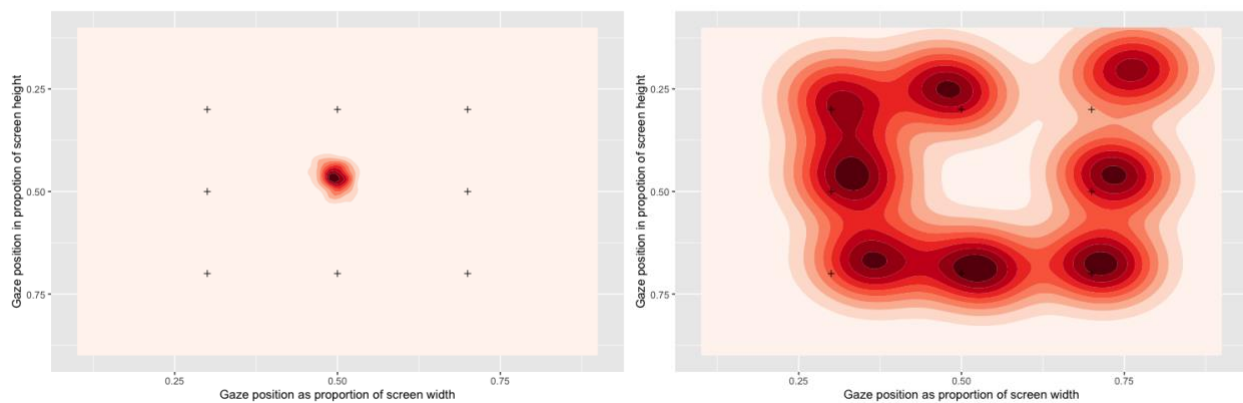
Spatial Precision

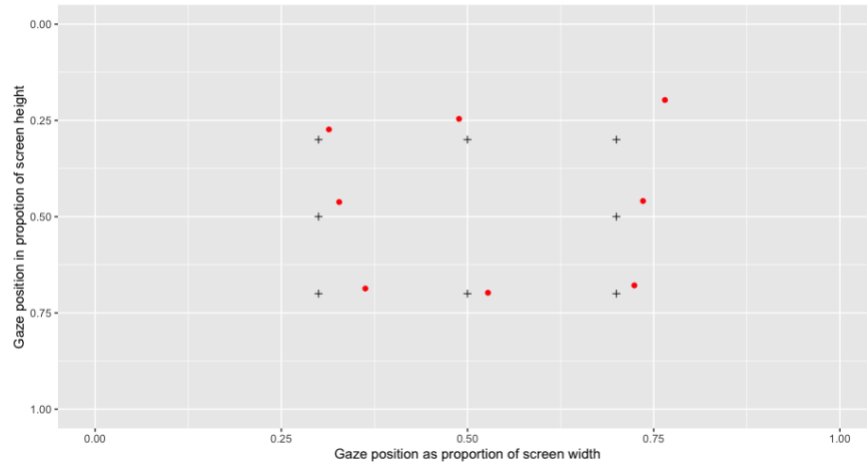
To determine the distribution of gaze points around the target locations, we calculated the average gaze coordinates within participants in 20 *ms* time bins at each target location.

Coordinates in the initial 1000 *ms* of a trial are shown in Figure 6 on the left, and data from the final 1000 *ms* is shown on the right. A grand average center of fixation for each target location was calculated for the final 1000 *ms*. As shown in Figure 6 (bottom), the grand average *Gazer* coordinates were close to the actual target locations (plotted as black crosses), but they were reliably different ($t(41) = 2.61, p < 0.02$).

Figure 6

Average Gaze Location Predictions as Recorded by Gazer in the First 1000 ms (left) and Final 1000 ms (right) for Trials with Each Target Location. Average Gaze Locations separated by Target Position (bottom).





Note. The spatial distribution of the points is congruent with the area of interests across trials, namely the red target dot positions.

Experiment 1B: Exogenous Attention task with Eyelink 1000

The goal of Experiment 1B was to compare the spatial and temporal results of *Gazer* on the exogenous attention task to the results derived from lab-based, corneal reflection eye tracking system (Eyelink 1000).

Method

Participants

Twenty participants (number of females = 17) ranging in age from 17 to 22 ($m = 19.3$ years) took part in this study in exchange for course credit. Of the twenty participants, sixteen were successfully calibrated and generated a complete dataset. All participants had normal or corrected to normal vision with contacts. The experiment was approved by the Research Ethics of the University of Victoria.

Materials and Equipment

All participants were tested in the Different Minds Laboratory at the University of Victoria using the same software and hardware settings (Matlab 2021b, PopOS, Alienware R8 Gaming Desktop, 240Hz refresh rate). Eye movements were recorded using the in-lab Eyelink 1000. The Eyelink performed monocular eye tracking with a sampling frequency of 1000Hz using the pupil and corneal reflection method. The experiments were created using Matlab 2020a and were presented on a separate monitor placed behind the eye-tracker (24.5" monitor with 240Hz refresh rate). The distance between the screen and participant was 15 inches (38 cm). Participants' heads were placed in a chin rest for the duration of the experiment. The room was lit with standard fluorescent lighting, and outside light was blocked for the duration of the experiment.

Procedure

A nine-point method of calibration was employed where participants were asked to fixate on nine points on the screen while the Eyelink tracked their gaze on these. The system records any disparity between the calculated gaze position and the target locations and employs a gaze correction adjustment. In a final validation phase, participants fixate on the same nine points a second time with some gaze adjustment. Validation was considered sufficient if the calculated gaze positions were within 1-degree visual angle from the targets. After participants successfully completed the calibration phase, they proceeded to the attention task.

The attention task was identical to the task used in Experiment 1A. At the beginning of each trial, a gaze-contingent cross was presented and replaced by the target circle stimulus at one of eight possible locations. Participants were instructed to execute an eye movement to the target circle and to remain fixated until it disappeared (a static 1500 ms). The location of the target

circle was randomized across each of the five blocks of eight trials, for a total of 40 trials. Participants completed the task in approximately 15-20 minutes.

Data Analysis

Eyelink Data Preparation. The raw EDF files for each participant were converted into ASC text files using the provided Eyelink software and imported into a Python analysis file. The raw data was parsed into a longform dataframe of gaze locations (x-coordinate and y-coordinates) with timestamps and target position information. The data was formatted to be like the raw *Gazer* output. The location coordinates were converted from pixel units to screen proportion units. For example, an x-coordinate would be converted using $\text{x-coordinate (pixels)} / \text{x-axis total pixels}$. The procedure for the Eyelink calculations of distance to target follow the same procedure outlined in Experiment 1A for the *Gazer* system.

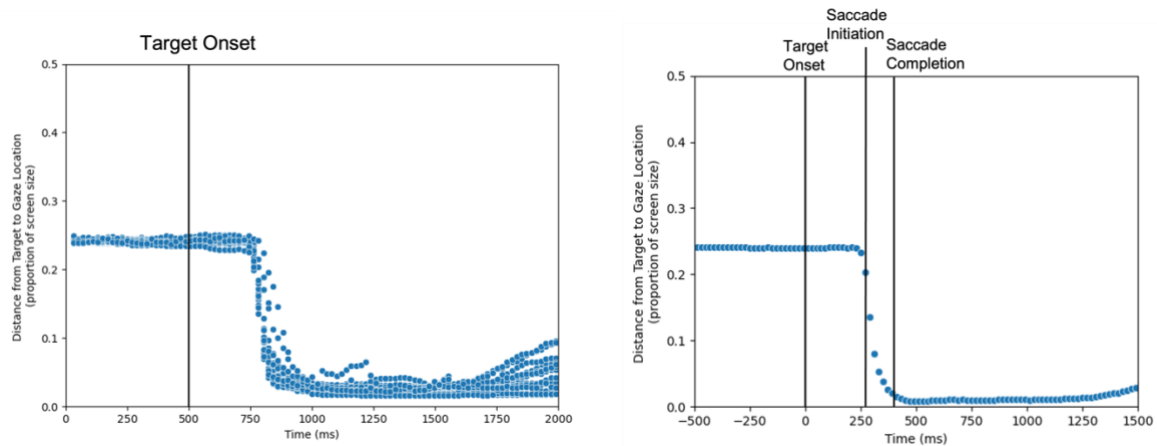
Results

Temporal Precision

The Euclidean distance between the red target dot, and the average gaze location across participants are visualized into a scatter plot. Figure 7 demonstrates two fixations (zero slope) were separated by an eye movement occurring towards the target (negative slope).

Figure 7

The Distance to Target Recordings for the Eyelink for Individual Participants (left) and Aggregated Across All Participants (right).



Note. The onset of the saccade towards the target occurs approximately 300 *ms* after target onset.

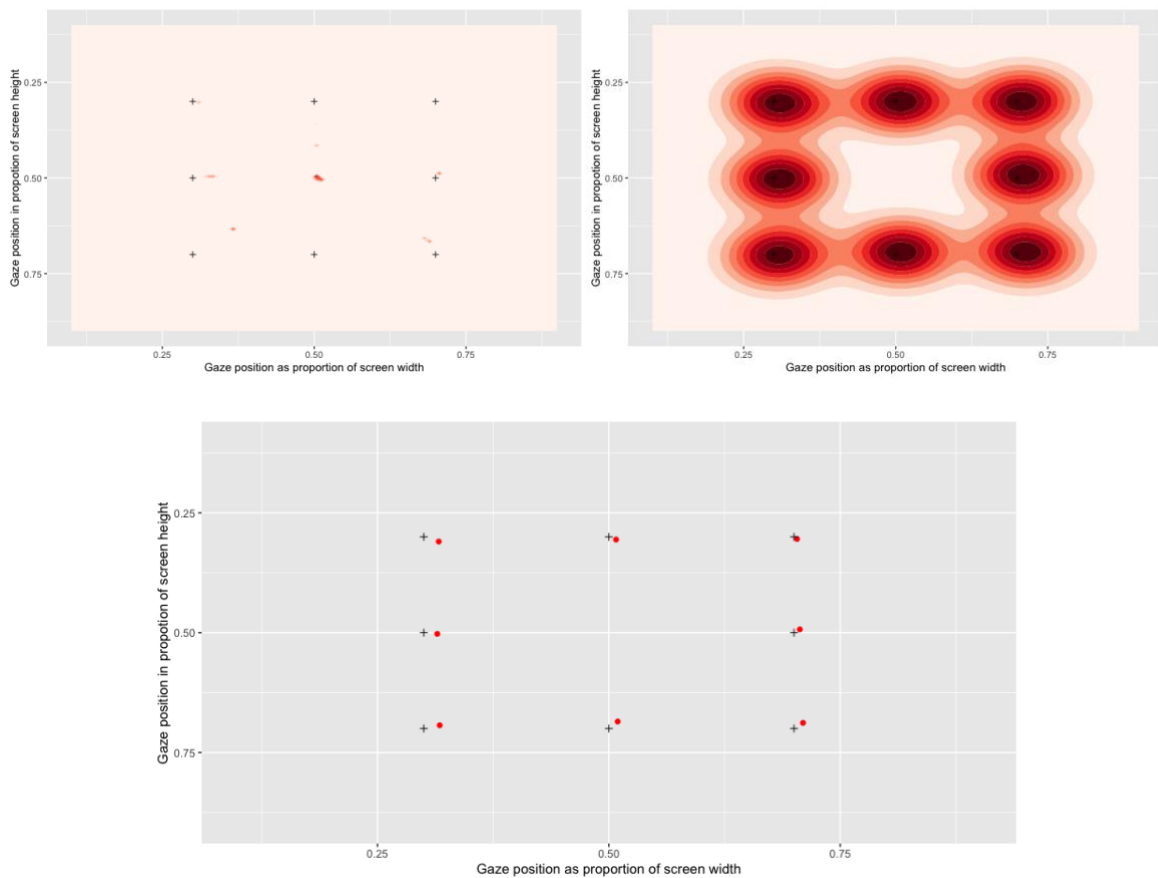
The onset and completion of the eye movement towards the target occurred an average of 291 *ms* and 345 *ms* after target onset. The time window of the saccade, calculated as the time between the onset and completion of the eye movement, was therefore approximately 53 *ms*.

Spatial Precision

The average locations of gaze within participant, in 20 *ms* time bins and separated by target location, are visualized in Figure 8. Locations measured in the initial 1000 *ms* of trials are shown on the left, and data from the final 1000 *ms* is shown on the right. Gaze coordinates recorded in the final 1000 *ms* were averaged between participants to calculate the grand average of fixation after onset, for each location (Figure 8, bottom). The Eyelink grand average centers of fixation were reliably different from the actual target locations ($t(41) = 2.58, p < .02$).

Figure 8

Average Gaze Location Predictions as Recorded by Eyelink in the First 1000 ms (left) and Second 1000 ms (right) for Trials By Target Location. Average Gaze Locations By Location in the Final 1000 ms (bottom).



Note. The spatial distribution of the points is congruent with the area of interest for each trial, namely the red target dot positions, indicated here using black crosses.

Comparing Experiments 1a and 1b Results

The main goal of this study was to compare the performance of the laptop *Gazer* system to the lab-based Eyelink system with respect to its spatial precision and temporal accuracy.

Spatial Precision of Gazer and Eyelink

As shown in Figures 5 and 7, both the *Gazer* and Eyelink systems registered the participant's saccade to the onset of the peripheral target stimulus. Whereas *Eyelink* was within 1% of the screen from the target, the *Gazer* system was less precise and within 5% of the screen. The absolute target coordinate and grand average gaze centers by target location are summarized in Table 1. *Eyelink* and *Gazer* coordinates were not reliably different from each other ($t(78) = 0.02, p = 0.99$), and showed a small but reliable difference from the actual target locations ($p < .02$).

Table 1

Absolute Coordinates of the Eight Red Dot Targets as Compared to the Calculated Clusters of Gaze Recorded by both Gazer and the Eyelink 1000. Clusters were Calculated as the Average Gaze Locations in the Final 1000ms of Trials. Coordinates Locations are Reported using Screen Proportion as a Metric.

Target Label	Measurement Device		
	Actual Coordinate (x,y)	Gazer (x,y)	Eyelink (x,y)
Target 1 (top left)	0.30,0.30	0.31,0.28	0.32, 0.31
Target 2 (top middle)	0.50,0.30	0.49, 0.25	0.51, 0.31
Target 3 (top right)	0.70,0.30	0.77, 0.20	0.70, 0.30
Target 4 (middle right)	0.70, 0.50	0.74, 0.46	0.71, 0.50
Target 5 (bottom right)	0.70, 0.70	0.72, 0.68	0.71, 0.69
Target 6 (bottom middle)	0.50, 0.70	0.53, 0.70	0.51, 0.69

Target 7 (bottom left)	0.30, 0.70	0.36, 0.69	0.32, 0.69
Target 8 (middle left)	0.30, 0.50	0.33, 0.46	0.32, 0.50

Temporal Precision of Gazer and Eyelink

The temporal precision of the systems was compared using the recorded onsets for toward-target saccades. Onsets were calculated as the first shift from a zero slope to non-zero slope in the distance to target data graphed in Figures 5 and 7. The Eyelink recorded the saccade onset on average approximately 291 *ms* after the target, whereas the recorded *Gazer* saccade onset occurred on average 370 *ms* after target onset. The temporal difference of 80 *ms* between the two systems was reliable, $p < .01$. The timing between saccade onset and completion were approximately 53 *ms* and 192 *ms* for Eyelink and *Gazer*, respectively. Hence, the time window of the Eyelink eye movement was narrower for the same spatial distance, as shown in Figures 5 and 7.

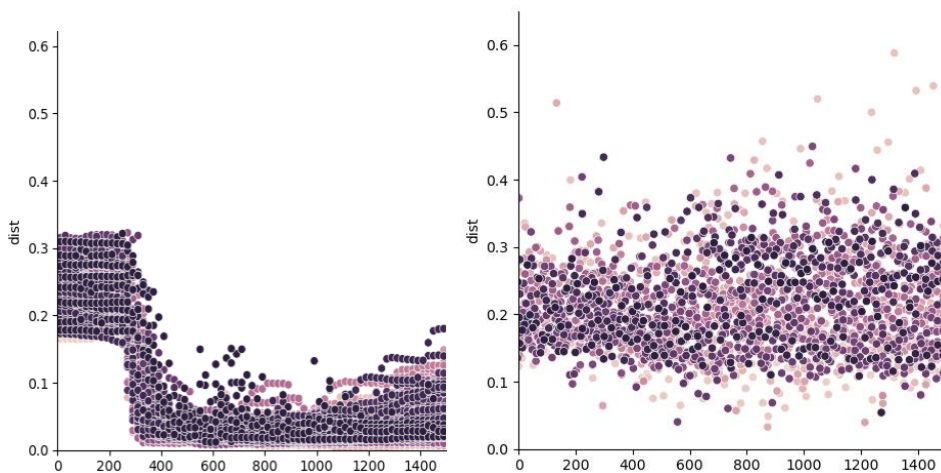
Individual Differences in Recording Eye Movements by Eyelink and Gazer

The onset for saccades to the target by Eyelink was 291 *ms* (s.d. =13.17) compared to saccade onset of 370 *ms* (s.d. = 121.99) for *Gazer*. Variance of the onsets was reliably greater for the *Gazer* data, as assessed by a Levene's test for homogeneity of variance ($p < .001$). Similarly, the variance of the recorded completion of saccades to the target (Eyelink: $m = 345$ *ms*, *Gazer*: $m = 562$ *ms*) was reliably greater for *Gazer* ($p < .001$). Hence, the variability of the data's precision, as shown by the standard deviation for the saccade onset and completion, is much greater across the *Gazer* participants.

To further illustrate this point, the distance to target points are plotted for all participants as recorded by the Eyelink (left) and *Gazer* (right) systems in Figure 9. Without aggregation across or within participants, the measurement noise obscures any gaze pattern in the *Gazer* data. In comparison, the gaze pattern amongst the Eyelink participants demonstrates little interindividual variability.

Figure 9

Distance to Target Measures for All Participants Recorded by Eyelink (left) and Gazer (right).



Individual differences within spatial precision are shown in Figures 6 (right) and 8 (right).

Discussion

The purpose of the current paper is to assess the viability of the *Gazer* web-based eye tracking system for attentional research. To do so, we compare the temporal and spatial precision of the *Gazer* system to the lab based Eyelink 1000. In addition, we want to evaluate the robustness of data produced by the remote *Gazer* system across different testing environments.

The performance of the two systems was assessed in a standard attentional cuing paradigm where a gaze-contingent fixation cross appeared for interval of 500 *ms*, followed by a target presented at one of eight spatial locations. Participants were instructed to make an eye movement to the target and maintain their fixation on the target for 1500 *ms*. To evaluate the temporal properties of both systems, we measured the timing difference between *Gazer* and Eyelink systems to detect the onset of a saccade to the target after the appearance of the red dot target. The saccade onset for *Gazer* was 370 *ms* (*s.d.* = 122) compared to Eyelink's onset of 292 *ms* (*s.d.* = 13). The 80 *ms* difference between the two systems was likely due to the slower sampling rate of *Gazer* (i.e., 30Hz) compared to faster sampling rate of the Eyelink system (1000 Hz).

As revealed by aggregate fixation points (Figures 6 and 8), the eye movements of both systems were accurately centered on the specified target locations. Although the fixations of the *Gazer* system were more dispersed (Figure 6) compared to the Eyelink's more densely clustered eye fixations (Figure 8), the identified centers from the target clusters were not reliably different across systems. In summary, the results examining the spatial properties of the *Gazer* system indicate that the web-based eye tracking data was similar to the data obtained with the laboratory-based Eyelink system. To summarize, the web-based *Gazer* produced research quality saccade detection data that was comparable in its spatial and temporal precision to a laboratory-based system.

A second goal of the study was to evaluate the robustness of the *Gazer* system to determine how well the system performed with a large sample of participants who completed the experiment under different testing environments with a variety of laptop devices. Across participants and testing conditions, the *Gazer* program performed well. In our study, 97% of the

participants (41 of 42 online participants) completed the self-calibration procedure and attention task and submitted valid data sets for analysis. The high success rate of *Gazer* was superior to the 33% completion rate of the participants (32 of 84 online participants) in the original *Webgazer* study (Simmelmann & Weigelt, 2018). A distinguishing factor of the calibration in the current study is the implementation of the “Point-and-Click” procedure that allowed for participants to refine the calibration of the system with their mouse. The original study by Semmelmann & Weigelt implemented a randomized method of calibration that relied on the default data of the face mesh, without any refinement. The authors also note calibration was approximately 50% of experimental time. In comparison, the “Point-and-Click” method used in the current study calibrated participants quicker and more effectively (*i.e.*, a greater percentage of participants proceeded past calibration). Indeed, the completion rate for *Gazer* also surpassed the in-lab Eyelink rate of 85% (*i.e.*, 16 of 20 participants), even though the in-lab participants were guided by a research assistant. Hence, the self-calibration procedure used in *Gazer* program was more efficient compared to the more tedious and time-consuming method applied in the *Webgazer* program (Simmelmann & Weigelt, 2018) and extensive calibration procedure used for the Eyelink’s system.

In the final evaluation, the researcher’s consideration of *Gazer* as an eye tracking system will depend on the scientific question under study and the practical constraints of the researcher. In terms of the research question, our study demonstrates that *Gazer* has the spatial precision to reliably discriminate a minimum of nine spatial regions (*i.e.*, the center fixation cross and eight red dot target locations). *Gazer*’s 30Hz sampling rate provided sufficient temporal data to record abrupt eye movements (*i.e.*, saccadic eye movements) and previous studies (Simmelmann & Weigelt, 2018) have also shown that the platform is capable of tracking moving objects (*i.e.*,

pursuit eye movements). Although *Gazer* is sufficiently robust to conduct basic eye movement experiments, the system might not be appropriate for tasks that demand high speed temporal updating such as reading (Rayner, 2022) or recognition of dynamic expressions (Liu *et al.*, 2014).

Despite these performance trade-offs, the web-based *Gazer* system affords the researcher many advantages over traditional in-lab, eye tracking methods. First, the pace of data collection is accelerated because participants can complete online experiments at their convenience. Eye tracking studies that might have taken weeks or even months to complete in the lab can be finished in three or four days (or less) over the web. Second, given the availability of web-based experiments, the researchers will have better access to populations (*e.g.*, older adults, young children) who might find it difficult to participate in an in-lab eye tracking experiment, or are at risk of experiencing stereotype threat (Spencer *et al.*, 2016). Third, because web-based eye tracking experiments are self-administered, the time and cost required for training research assistants are significantly reduced or even eliminated. Most importantly, the *Gazer* system is available as an open-source experimental package to be used in conjunction with the online experimental platform jsPsych¹. As with other open-source software, researchers have the capability to implement, modify and extend the *Gazer* code to meet the specific demands of their research program.

Beyond research applications, systems such as *Gazer* may be suitable for industry and educational purposes. Individuals interested in how visual interest is captured on websites or by advertisements may be able to use this eye tracking technology to enhance marketing and sales techniques. *Gazer* may also be useful for participants as an alternative to the keyboard and

¹ The *Gazer* system is freely available for downloading at <https://github.com/amyvanwell>.

mouse for computer interaction. Instead of selecting a button with a mouse, participants may be able to interact with it through moving their eyes and focusing their gaze on the button's location.

In summary, the primary goal of the *Gazer* project was to make the program easily available for researchers in the vision sciences community. By increasing the accessibility of web-based eye tracking technology, we can begin to see an even greater expansion of its applications and hone our understanding of the research questions that it is suited to answer. We believe that the *Gazer* system is a viable eye tracking tool to address broad range of research questions in cognitive psychology related to perception, attention, memory, and decision-making.

References

- Alonso, V. E., Enríquez-Caldera, R., & Sucar, L. E. (2017). Foveation: an alternative method to simultaneously preserve privacy and information in face images. *Journal of Electronic Imaging*, 26(2), 023015.
- Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2021). *MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture*.
<https://doi.org/10.31234/osf.io/rsdwg>
- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, 28(3), 241–248.
- de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of Experimental Psychology. Human Perception and Performance*, 14(2), 188–202.
- Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: an eye-tracking analysis. *Frontiers in Psychology*, 3, 335.
- Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, 6(1). <https://doi.org/10.16910/jemr.6.1.3>
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology. Human Perception and Performance*, 16(4), 802–811.

- Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning*, 33(3), 193–201.
- Jenke, L., Bansak, K., Hainmueller, J., & Hangartner, D. (2021). Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association*, 29(1), 75–101.
- Liu, M., Shan, S., Wang, R., & Chen, X. (2014, June). Learning Expressionlets on Spatio-temporal Manifold for Dynamic Facial Expression Recognition. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA.
<https://doi.org/10.1109/cvpr.2014.226>
- Papoutsaki, A. (2015). Scalable Webcam Eye Tracking by Learning from User Interactions. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 219–222.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Rayner, K. (2022). Foveal and Parafoveal Cues in Reading. In *Attention and Performance VII* (pp. 149–161). Routledge.
- Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2), 451–465.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype Threat. *Annual Review of Psychology*, 67, 415–437.

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 746–759.

Wu, E. X. W., Laeng, B., & Magnussen, S. (2012). Through the eyes of the own-race bias: eye-tracking and pupillometry during face recognition. *Social Neuroscience*, 7(2), 202–216.

**Chapter 3: Where's Waldo? Exploring Gaze Strategy in a Visual Search Task Online and
In-Person**

Amy vanWell¹ & James W. Tanaka¹

1. Department of Psychology, University of Victoria

Abstract

In this study, we employed a ‘Where’s Waldo’ visual search task to compare the eye movement behaviours of participants in-person and online. Participants were presented with a “Where’s Waldo” scene and their task was to find the ‘Waldo’ target in the scene. They pressed the spacebar to indicate when they had located the target then used the mouse to click on the target’s location. Eye-tracking data was recorded using the online Gazer ($n = 106$) and the in-person Eyelink 1000 ($n = 18$). The online participants were equally accurate at identifying the target’s location (online: 99% accuracy v. in-lab: 96%) but were significantly slower to press the space bar to indicate target detection (online: 7210 *ms* v. in-lab: 6460 *ms*). However, analysis of eye tracking data showed an opposite pattern of eye movement behaviours where the online participant’s first fixations on the target location occurred 3000 *ms* sooner than the first fixations of the in-lab participants (online: 3764 *ms* v. in-lab: 6407 *ms*). Therefore, online participants were fixating on the target more quickly, but were slower to indicate detection. A region of Interest (ROI) analysis indicated that in-lab participants averaged 1.53 fixations on the area containing the Waldo target and online participants averaged 3.45 fixations, but this difference was not reliable. The in-person participants appeared to respond after a single fixation on the target, but the online participants made several saccades to and away from the target before responding. Online participants may be demonstrating a more conservative strategy in the visual search task by only providing the detection response after confirming target presence in multiple passes. Overall, we recorded compelling eye-tracking and behavioural data both in-person and online and provide evidence that remote participants may use an altered gaze strategy in a visual search task.

Previous research demonstrates behavioural data collected remotely shows similar precision to data collected in person for the same experiment (de Leeuw 2015; Kochari 2019; Eerola *et al.* 2021; Anwyl-Irvine *et al.* 2020). However, since the development of remote testing programs, the demand for new data measures historically collected only in-lab has emerged. At the University of Victoria, we have set out to expand remote data collection for psychological experimentation with the development of a web-based eye tracking system, *Gazer*. The *Gazer* system uses the open-source *Webgazer* (Papoutsaki, 2015) software to make real time predictions of screen-based gaze locations. The raw video feed from remote participant's video feed is accessed to acquire "whole head" information from the orientation of their facial features.

Eye tracking is a common data collection method used within cognitive psychology. How a participant views to a visual stimulus (*i.e.*, where a participant looks, and in what order) is informative for understanding visual processing, and other cognitive processes including memory (Zacks, 2020; Watson *et al.*, 2019) and attention (Armstrong & Olatunji 2012; Shagass *et al.*, 1976; Chita-Tegmark, 2016). Eye movements are commonly divided into "exogenous" and "endogenous" types based on the source of attention driving the eye movement. The simple and rapid exogeneous eye movement is generated from the sudden onset of a stimulus and can be used as marker for some of the fastest eye movements made by participants. In contrast, endogenous eye movements are less direct and driven top-down by the participant, usually in line with a task or goal. In a previous experiment, we compared the "whole head" eye tracking method used by *Gazer* to the established in-lab corneal reflection Eyelink eye tracking system and concluded it had comparable spatial and temporal precision for an exogeneous eye movement. In the current study, we aim to expand our comparison between *Gazer* and Eyelink

beyond exogenous eye movements to compare recording for endogenous eye movements using a visual search task based on the classic Where's Waldo game.

EXPERIMENT 1A: WHERE'S WALDO WITH GAZER

In Experiment 1A, we used the *Gazer* eye tracking system to record eye movements in a visual search task. The goal of the experiment was to record behavioural performance for locating Waldo, and eye movement data that may indicate potential search strategies.

Method

Participants

A total of 106 participants (# females = 96, $m = 21.93$ years) were recruited through UVic's SONA system. The participants had normal or corrected-to-normal vision with contacts. Participation was limited to those with access to a personal laptop with a Chrome browser.

Materials and Equipment

Experimentation was performed remotely using personal laptops. Participants used various hardware systems; with the limitation they use a standard personal laptop. The experiment was hosted on Heroku and accessed by participants using a URL link. The data was saved to a MongoDB database hosted on DigitalOcean.

Procedure

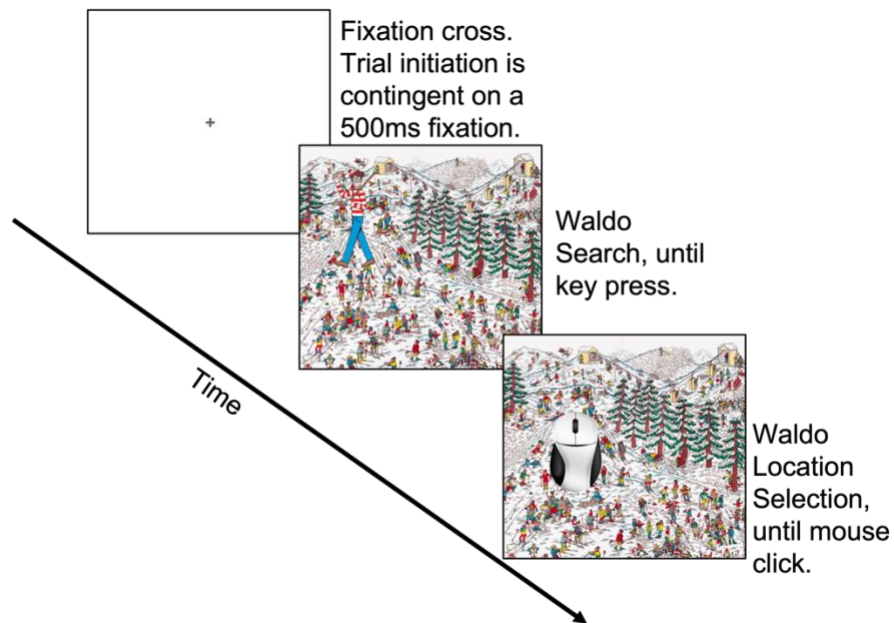
Eye tracking was performed using the web-based *Gazer* system. Prior to the experimental trials, participants performed a 9-point "Point-and-Click" self-calibration until they reached a validation threshold of 70% accuracy. The video feed of the laptop was continuously accessed

during experimentation to provide real-time predictions of screen-based gaze locations at a frequency of 30 *Hz*.

Before each trial, participants were shown a gaze-contingent fixation cross. They proceeded to the Where's Waldo search scene once 500 *ms* of consecutive fixation was recorded on the cross. Immediately after, the background scene appeared (*i.e.*, trials started) with the Waldo target superimposed on top (Figure 1, slide 2) and visual search for the Waldo target was initiated. The mouse was not visible during the search portion of the task. Participants were provided with 20 *seconds* to locate Waldo and press the [space] key to indicate he was found. The location of the target was one of eight possible corner locations, each placed 30% of the screen's x-axis and y-axis away from the screen's edge. After the [space] key was pressed, the Waldo target was removed from the scene and participants were asked to use their mouse to click on the target's location. The location of the target Waldo was randomized across each of the eight blocks of eight trials, for a total of 64 trials. A new background was introduced in each block, in random order between participants.

Figure 1

A Sample Sequence of a Where's Waldo Trial.



Note. Trials are initiated after 500 *ms* of consecutive fixation on a centered cross. Immediately after, the Waldo search scene is presented, with the Waldo target superimposed at one of eight random locations. Participants have 20 *s* to end the search by pressing the [space] key at which point Waldo is removed from the scene and the participant must use their mouse to click on the target's previous location.

Data Analysis

Analysis was performed on accuracy (trials were classified as correct if the participant clicked on the correct location with their mouse) and reaction time (RT) for correct-only trials separately as RT to press the [space] key and to perform the mouse click. An analysis of variance was performed to analyze how the behavioural measures varied across background.

A region of interest (ROI) analysis was performed using eye tracking data recorded by *Gazer*. The ROI was defined as the 162 *pixel* by 50 *pixel* rectangular space that contained the Waldo target. Using the raw gaze points predicted by *Gazer*, the average onset of the first fixation in this region was calculated, and the average total number of fixations per trial. Number of fixations per trial was calculated using a dispersion threshold algorithm introduced by Salvucci & Goldberg (2000). A separate “distance to target” measure was calculated as the Euclidean distance between the average gaze points recorded by *Gazer* in 20 *ms* time bins across participants, and the location of the target Waldo. Distance was reported as a proportion of the screen size to accommodate different screen sizes between participants. Data was time-locked to the [space] press (*i.e.*, the data point at 1000 *ms* represent the average distance to target recorded exactly 1000 *ms* before the key press).

Results

Before investigating patterns in eye tracking measures, we probed reaction time (separately as the time to press the [space] bar and to make the mouse response) and accuracy for the mouse response.

Behavioural Results

Descriptive measures for reaction time ([space] key and mouse click), and accuracy are summarized in Table 1. Participants made few errors on the task (error rate = 4%) and completed trials in approximately 8410 *ms* (*s.d.* = 7765).

Table 1

Behavioural (Reaction Time and Accuracy) Measures Recorded During a Web-based Where's Waldo Experiment.

	Reaction Time ([space] key)	Reaction Time (mouse click)	Accuracy (mouse click)
Mean	6460 <i>ms</i>	1950 <i>ms</i>	96 %
Standard Dev	4473 <i>ms</i>	3292 <i>ms</i>	18 %

Note. Reaction time reported for correct trials only.

Background Effects. Remote participants showed reliable differences in the time required to find the target (press the [space] key) across backgrounds ($F(103) = 22.04, p < .001$). Participants were fastest (4927 *ms*) to click the [space] key for background 8 (7772 *ms*, shown in Figure 2, bottom), and slowest for background 2 (Figure 2, top). Interstimulus perceptual variance between backgrounds 2 and 8 was probed (Bainbridge and Oliva 2015). Background 2 demonstrated reliably greater spatial frequency ($p < .001$) and lower lightness ($p < .001$). A colour histogram indicated significantly different RGB values between the pictures ($p < .001$). Accuracy was not different across backgrounds ($p = .89$).

Figure 2

Two Example Backgrounds Implemented in the Where's Waldo Task.



Note. Remote participants were slowest to find the target for background 2 (top) and fastest for background 8 (bottom). Waldo was randomly superimposed on the images at one of eight locations.

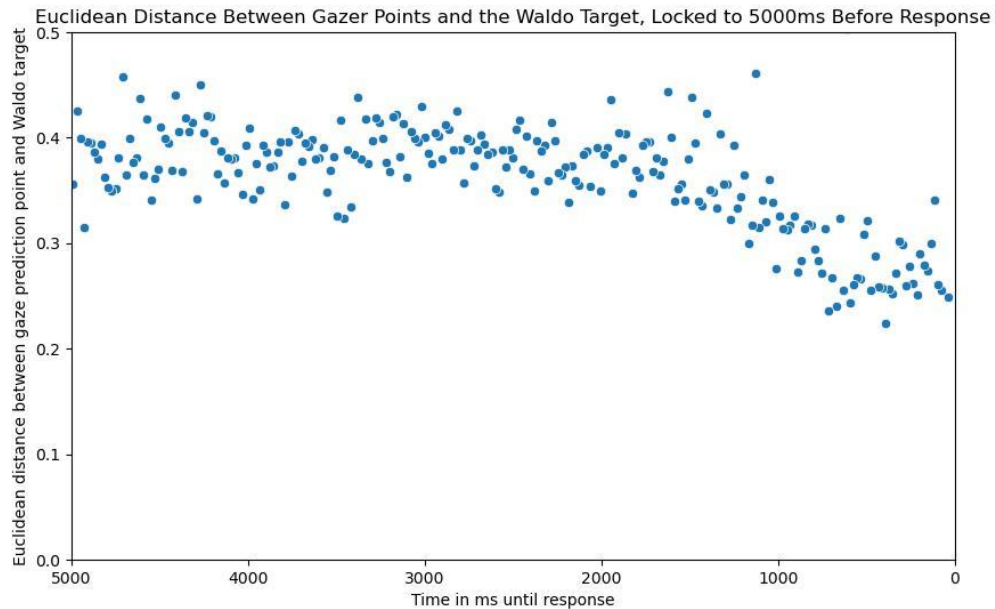
Eye Tracking Measures

Region of Interest Analysis. The first fixation made to the defined region of interest (rectangle surrounding the target Waldo's location) was made approximately 3734 *ms* (*s.d.* = 3029 *ms*) after the scene appeared. Participants made approximately 3.45 fixations (*s.d.* = 1.68) on the target per trial.

Distance to Target Time-Locked to the Key Press. The distance between participant's average gaze locations and the Waldo target are shown for the final 5000 *ms* of trials in Figure 3. A sharp eye movement is made towards the target's location starting approximately 2000 *ms* before participants press the [space] key to indicate they have located Waldo.

Figure 3

Euclidean Distance Between Participants Average Gaze Location Recorded by Gazer in 20 ms Time Bins and the Target Waldo in the 5000 ms Immediately Before the [space] Key Press was Made.



EXPERIMENT 1B: WHERE'S WALDO WITH EYELINK

The goal of Experiment 1B was to compare the eye tracking and behavioural results of the online visual search task to results recorded in-lab using the Eyelink 1000.

Method

Participants

A total of 18 participants (# females = 13, $m = 24.24$ years) were recruited through UVic's SONA system. The participants had normal or corrected-to-normal vision with contacts.

Materials and Equipment

All participants were tested in the Different Minds Laboratory at the University of Victoria using the same software and hardware settings (Matlab 2021b, PopOS, Alienware R8 Gaming Desktop, 240Hz refresh rate). Eye movements were recorded using the in-lab Eyelink 1000. The Eyelink performed monocular eye tracking with a sampling frequency of 1000Hz using the pupil and corneal reflection method. The experiments were created using Matlab 2020a and presented on a separate monitor placed behind the eye-tracker (24.5" monitor with 240Hz refresh rate). The distance between the screen and participant was 39 centimeters. Participants' heads were placed in a chin rest for the duration of the experiment. The room was lit with standard fluorescent lighting, and outside light was blocked for the duration of the experiment.

Procedure

Eye-tracking was performed using the in-lab Eyelink 1000. Prior to experimentation, participants underwent a 9-point calibration process, guided by a research assistant. The visual search task was identical to the task used in Experiment 2A. At the beginning of each trial, a gaze-contingent cross was presented and replaced by Where's Waldo scene with the Waldo target superimposed at one of eight possible locations. The location of the target Waldo was randomized across each of the eight blocks of eight trials, for a total of 64 trials. A new background was introduced in each block, in random order between participants. Participants completed the task in approximately 20-25 minutes.

Data Analysis

Analysis was performed on accuracy and reaction time (RT) for correct-only trials separately for RT to press the [space] (*i.e.*, indicate the target was found) key and perform the mouse click (*i.e.*, indicate the target's location). An analysis of variance was performed to determine if the behavioural measures varied across background.

A region of interest (ROI) analysis was performed using the DataView application. The ROI was defined as the rectangular screen area (152 *pixels* by 50 *pixels*) containing the Waldo target. A report was generated for first fixation time, and number of fixations on the target, per trial. The procedure for calculating distance to target locked to the [space] key followed the same steps outlined in experiment 1A.

Results

Behavioural Results

Descriptive measures for reaction time ([space] key and mouse click) and accuracy are summarized in Table 2. Participants showed a large amount of variability for the initial [space] key press (12237 *ms*) but consistently fast mouse clicks on the target. Accuracy for the task was near ceiling at 99%.

Table 2

Behavioural (Reaction Time and Accuracy) Measure for an In-lab Where's Waldo Experiment.

	Reaction Time ([space] key)	Reaction Time (mouse click)	Accuracy (mouse click)
Mean	7210 <i>ms</i>	1567 <i>ms</i>	99 %
Standard Dev	12237 <i>ms</i>	379 <i>ms</i>	12 %

Note. Reaction Time calculated for correct trials only.

Background Effects. The in-person participants showed reliable differences in reaction time across different backgrounds ($F(123) = 14.52, p < .001$). In-lab performance was slowest for

background 2 (15100 *ms*, Figure 2, top), and fastest for background 8 (2620 *ms*, Figure 2, bottom). As reported in Experiment 1a, the low-level features of spatial frequency, colour, and lightness were significantly different ($p < .001$) between backgrounds 2 and 8. Accuracy was not reliably different across background ($p = .16$).

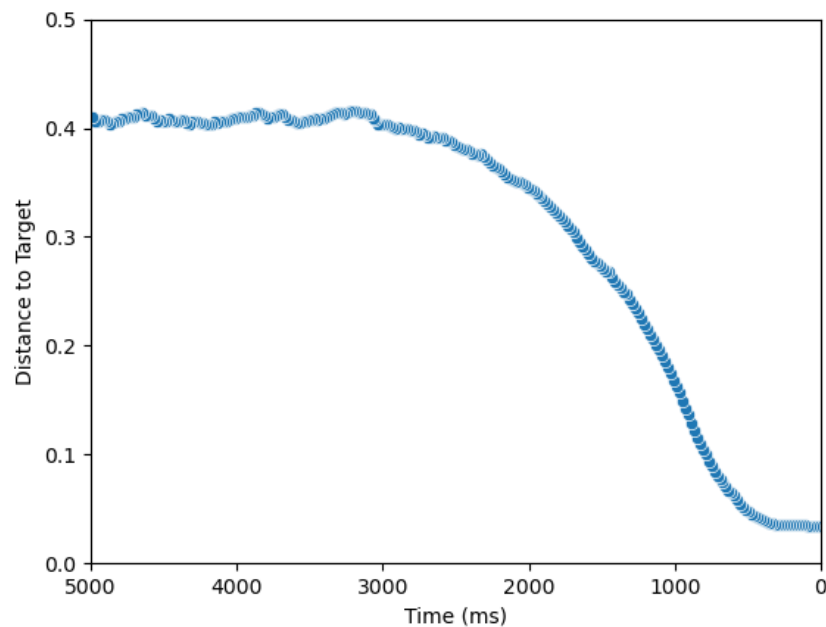
Eye Tracking Measures

Region of Interest Analysis. The first fixation made to the defined interest area (rectangle surrounding the target Waldo's location) was made approximately 6407 *ms* (*s.d.* = 10068 *ms*) after the scene appeared. On average, participants made 1.53 fixations (*s.d.* = 0.99) on the target in each trial.

Distance to Target Time-Locked to the Key Press. Distance to target for the 5000 *ms* of the trial immediately prior to the [space] key press is shown in Figure 4. Participant's gaze is shown to shift towards the target at the 3000 *ms* mark, with a fixation being performed in the 500 *ms* interval before the key press.

Figure 4

Euclidean Distance Between Participants Average Gaze Location Recorded by Eyelink in 20 ms Time Bins and the Target Waldo in the 5000 ms Immediately Before the [space] Key Press was Made.



COMPARISON OF RESULTS BETWEEN EXPERIMENTS 1A AND 1B

The goal of the current experiment is to evaluate differences between behavioural and eye tracking data collected for an identical experiment online and in-lab. An initial analysis of the behavioural data indicated that participant's performance was comparable between the testing environments.

Behavioural Results

Reaction Time

A one-way ANOVA by testing condition (remote versus in-person) indicated reliably longer reaction time was recorded for the initial space click during the in-lab experiment (7210

ms) compared to the online experiment (6460 *ms*, $F(122) = 18786$, $p < .001$). In contrast, participants performed the mouse click faster in-person (1567 *ms*) than online (1950 *ms*, $F(122) = 13021$, $p < 001$). The total length of trials was comparable (remote: 8410 *ms* versus in-lab: 8776 *ms*), but reliably different ($F(122) = 13021$, $p < 001$). Therefore, participants performed the visual search and location identification task at different rates, but the speed across both tasks was similar. The variability of RT presented an opposite pattern between the online and in-person datasets. In-lab, participants were inconsistent in their speed to make the [space] key response but showed little variability in the speed of the mouse click. In contrast, remote participants were consistent in their speed to indicate the target was found, but not for the mouse response. Overall, participants performed the complete trials in similar time across testing conditions.

The speed to find the Waldo target was reliably different across backgrounds, however the pattern of differences was identical for in-lab and remote participants. Participants in both testing conditions were slowest to identify Waldo targets shown on background 2 (Figure 1, top) and fastest for background 8 (Figure 2, bottom). An analysis of interstimulus perceptual variance indicated participants were fastest for the background that had greater lightness and higher spatial frequency (Bainbridge and Oliva 2015). Hence, the low-level properties influenced search performance similarly in both experiments, despite the variable hardware (*i.e.*, screen, screen settings) between conditions and across participants in online experimentation.

Accuracy

Accuracy was not reliably different between testing conditions ($p = 0.10$). Both online and in-person, participants were at ceiling for the task (online: 97% versus in-lab: 99%). Performance did not vary by background in either testing environment ($p = .89$). Therefore, the

interstimulus perceptual variance impacted the speed of participant's search to find the target but did not inhibit participants finding the target.

Eye Tracking Measures

Region of Interest Analysis

The onset of the first on-target fixations were reliably faster as recorded Gazer (3734 *ms*) compared to Eyelink (6407 *ms*, $F(120) = 77.81, p < .001$). Although not reliable ($F(121) = 0.02, p = .89$), remote participants did make more fixations (3.45 fixations) than their in-person counterparts (1.53 fixations) on the target, indicating they may make more returns to the target, despite fixating on it earlier.

Distance to Target Locked to the Key Press

An eye movement towards the target was present in eye tracking data recorded online and in-person. However, the trajectory of the eye movement was steeper for the remote dataset. *Gazer* recorded a sharp shift in location was present approximately 2000 *ms* before remote participants pressed the [space] key (Figure 3). The Eyelink 1000 recorded a slower shift by in-lab participants towards the target starting approximately 3000 *ms* before the key press (Figure 4). In both datasets, a fixation towards the target location was present in the last 500 *ms* of trials. However, the fixation recorded by Eyelink was closer to the target than the location recorded by *Gazer* (Figure 3).

Discussion

In a previous study (vanWell & Tanaka, 2023), we confirmed the viability of *Gazer* to record exogenous eye movements. The aim of the current study was to extend experimentation to

account for endogenous eye movements by directly comparing both behavioural and endogenous eye movement performance online and in-person.

An analysis of RT and accuracy indicated similar performance across testing conditions. Participants total RT to complete trials was comparable online and in-person, and there were no reliable differences for accuracy. Although there were influences from the low-level properties of the chosen Waldo backgrounds, the pattern of performance across backgrounds was identical for our remote and in-lab datasets. In other words, ISPV did reduce performance based on lightness and spatial frequency but did so in a consistent manner across screens used by participants. These results provide further evidence that remote testing produces comparable data compared to in-lab for behavioural measures.

Eye movements were compared based on fixations on the Waldo target. The onset of the first on-target fixation was recorded reliably earlier online than in-lab by approximately 3000 *ms* ($p < .001$). In a previous study, we found evidence for an 80 *ms* temporal lag for recording eye movements between Gazer and Eyelink (vanWell & Tanaka, 2023). Therefore, the current discrepancy is too large to be explained using a different eye tracking system, and instead may reflect a search strategy difference between remote and in-person participants. Previous studies in memory (Diamond *et al.*, 2020) have demonstrated strategy discrepancies between online and in-lab participants, and a previous study implementing Webgazer (Simmelman & Weigelt, 2018) found greater variability for data recorded remotely versus in-person. In the current study, no reliable difference was present between the total number of on-target fixations recorded between testing systems. An analysis of distance to target revealed that both systems captured an eye movement towards the target immediately before participants made a [space] response indicating

they had located the target, however the *Gazer* recorded a more direct movement compared to Eyelink.

A key element of the current paradigm is the introduction of the [space] key to indicate the target had been located. The key press enables comparison between the first on-target fixation as recorded by *Gazer* and Eyelink, and the participant's overt indication the target had been found. Remote participants demonstrated an approximate 2700 *ms* discrepancy between the initial fixation on the target and the key press, 1900ms more than their online counterparts (800 *ms*), providing clearer evidence for search strategy differences between testing conditions. However, although participants were tasked to press the [space] key as soon as they saw the Waldo target, there was an additional task beyond target detection. The target disappeared when the participants pressed the [space] key, therefore participants needed to encode his location before the key press to accurately click on his location with their mouse. Therefore, participants may have held off pressing the key until they had actively recorded his location in memory, potentially explaining the length of the time intervals between the first target fixation and the key press.

A further examination of the eye movement data revealed additional potential gaze strategy differences between testing conditions. Online participants found the target Waldo earlier and may have made more returns to his location, but the difference was not reliable. They also make a more direct eye movement to the target prior to pressing the [space] key. These findings indicate a more conservative strategy, such that online participants make multiple fixations on the target before ending the search. The online component of the task may reduce confidence in their responses, as they lose any response feedback after practice trials and have no research assistant present. As well, remote participants may have knowledge of the target's

location from additional previous fixations compared to their in-person counterparts, enabling the more direct final eye movement to his location. Accuracy did not differ across testing conditions; therefore, the potential conservative search strategy may have only influenced the speed of their response.

Conclusions

We demonstrate comparable recording for behavioural results and eye tracking measures online and in-person. The current work expands on previous results indicating sufficient precision by *Gazer* to record eye movements in an attentional cueing task (vanWell & Tanaka, 2023) to indicate its validity to record endogenous attentional strategies. In a follow-up experiment, we will further probe *Gazer's* ability to record gaze strategy by evaluating changes in fixations across multiple sessions of Where's Waldo.

References

- Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo K. Evershed. 2020. "Gorilla in Our Midst: An Online Behavioral Experiment Builder." *Behavior Research Methods* 52 (1): 388–407.
- Armstrong, Thomas, and Bunmi O. Olatunji. 2012. "Eye Tracking of Attention in the Affective Disorders: A Meta-Analytic Review and Synthesis." *Clinical Psychology Review* 32 (8): 704–23.
- Bainbridge, Wilma A., and Aude Oliva. 2015. "A Toolbox and Sample Object Perception Data for Equalization of Natural Images." *Data in Brief* 5 (December): 846–51.
- Chita-Tegmark, Meia. 2016. "Social Attention in ASD: A Review and Meta-Analysis of Eye-Tracking Studies." *Research in Developmental Disabilities* 48 (January): 79–93.
- Diamond, N. B., Abdi, H., & Levine, B. (2020). Different patterns of recollection for matched real-world and laboratory-based episodes in younger and older adults. *Cognition*, 202, 104309.
- Eerola, Tuomas, James Armitage, Nadine Lavan, and Sarah Knight. 2021. "Online Data Collection in Auditory Perception and Cognition Research: Recruitment, Testing, Data Quality and Ethical Considerations." *Auditory Perception & Cognition* 4 (3–4): 251–80.
- Kochari, Arnold R. 2019. "Conducting Web-Based Experiments for Numerical Cognition Research." *Journal of Cognition* 2 (1): 39.
- Leeuw, Joshua R. de. 2015. "JsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser." *Behavior Research Methods* 47 (1): 1–12.
- Papoutsaki, Alexandra. 2015. "Scalable Webcam Eye Tracking by Learning from User Interactions." In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on*

Human Factors in Computing Systems, 219–22. CHI EA '15. New York, NY, USA:
Association for Computing Machinery.

Shagass, C., R. A. Roemer, and M. Amadeo. 1976. "Eye-Tracking Performance and Engagement of Attention." *Archives of General Psychiatry* 33 (1): 121–25.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2), 451–465.

vanWell, A. & Tanaka, J.W. 2023. "Gazer: A New Tool for Web-based Eye Tracking". PsyArXiv.
<https://psyarxiv.com/2pckv>.

Watson, Poppy, Daniel Pearson, Michelle Chow, Jan Theeuwes, Reinout W. Wiers, Steven B. Most, and Mike E. Le Pelley. 2019. "Capture and Control: Working Memory Modulates Attentional Capture by Reward-Related Stimuli." *Psychological Science* 30 (8): 1174–85.

Zacks, Jeffrey M. 2020. "Event Perception and Memory." *Annual Review of Psychology* 71 (1): 165–91.

**Chapter 4: Evidence for Visual Search Performance Improvement through Perceptual
Training**

Amy vanWell¹ & James W. Tanaka¹

1. Department of Psychology, University of Victoria

Abstract

In the current study, we examined gaze strategy development and training effects in a repeated visual search task. Using the web-based *Gazer* system, we tracked the eye movements of participants remotely during a search paradigm modelled after the classic Where's Waldo game. In each trial, participants were presented with a Where's Waldo scene and asked to identify the location of the Waldo target. Once they located Waldo, participants terminated the search by pressing the space bar. After ending the search, they used their mouse to click on the target's location on-screen. During the visual search, gaze points were recorded at a sampling rate of 30Hz and timing was locked to the space bar response. Participants repeated the task across 3 sessions, with the same backgrounds repeating across sessions. At the level of individual differences, participant showed baseline (*i.e.*, no training) differences for total number of fixations made per trial. After each session, performance trended to be both behaviorally faster and make fewer fixations. Within each session, participants who were faster also made less fixations. Our results indicate participants became more efficient with their eye movements by making fewer and more focused movements but were not moving their eyes faster. A closer analysis of gaze strategy indicated participants may have executed a clockwise search strategy initiating at the top left corner of the screen. Overall, the results demonstrate *Gazer*'s capability to explore eye movements and make inferences about strategy and performance.

Within a complex visual search task (*e.g.*, looking for a blue square amongst red triangles), the number of fixations made in a single trial can be indicative of overall performance. Participants who use more efficient glancing and scanning, *i.e.*, make fewer fixations, tend to perform these tasks faster and more accurately (Williams *et al.*, 1997). Strong individual differences are evident even within baseline performance (*i.e.*, no training) for visual search (Joseph *et al.*, 2009). From this observation, studies have explored whether those with lower baseline performance can be trained to “catch up” to their more naturally performant peers. Research indicates perceptual training can improve task performance by reducing the number of fixations made in a single trial (Schuster *et al.*, 2013). The same performance improvement can be seen in the comparison of expert versus novice radiologists, who are trained to quickly locate a tumor target within a cancer screening (a real-life version of visual search). Expert radiologists scan screenings with greater accuracy and faster performance than novices and do so with fewer and more refined eye movements (Drew *et al.*, 2013). In the current task, we attempt to record similar training effects in eye movements within an experimental visual search paradigm.

Where’s Waldo Task

In the current study, we use a visual search task modeled after the popular Where’s Waldo game. Participants complete the same experiment on 3 separate occasions, approximately 24 hours apart (randomized order of trials in each session). In each trial, participants are shown a classic Where’s Waldo background (filled with intended visual distractors) and are asked to locate a Waldo character that has been superimposed onto the background randomly at one of 8 positions.

Figure 1

An Example of a Background Used in a Where's Waldo trial, with a Waldo Target Superimposed on the Top Left of the Background.



Note. Participants would press the [space] key once they located Waldo's location. Waldo is not superimposed to the size used in the experiment; he is enlarged for illustrative effects.

Research Questions

Based on previous research investigating differences between experts and novices in visual search tasks, we theorized performance for finding Waldo would become more efficient with each session. We capture performance efficiency using three different measures. Using reaction time, we investigated general behavioural changes across sessions. To evaluate eye movements, we calculated the average number of fixations made during trials and directly compare the resulting pattern to our behavioural results. We further evaluated the connection between behavioural and eye movement performance by calculating and investigating "fixation rate" as the number of fixations made per second. Beyond making fewer eye movements,

participants may be improving on an even more fine level, and making faster movements as well. We evaluate our data using multilevel modeling to determine how participants compare at baseline before training, and how their individual performance changes across sessions.

Method

Participants

A total of 36 (# females = 29, $m = 21.17$ years) remote datasets were collected, 28 were complete. The participants had normal or corrected-to-normal vision with contacts. Participation was limited to those with access to a personal laptop with a Chrome browser.

Materials and Equipment

The experiment was conducted remotely through participant's Chrome browsers. Participants used various hardware systems, with the limitation that they used a standard personal laptop. The experiment was hosted on Heroku and accessed by participants using a URL link. The data was saved to a MongoDB database hosted on DigitalOcean.

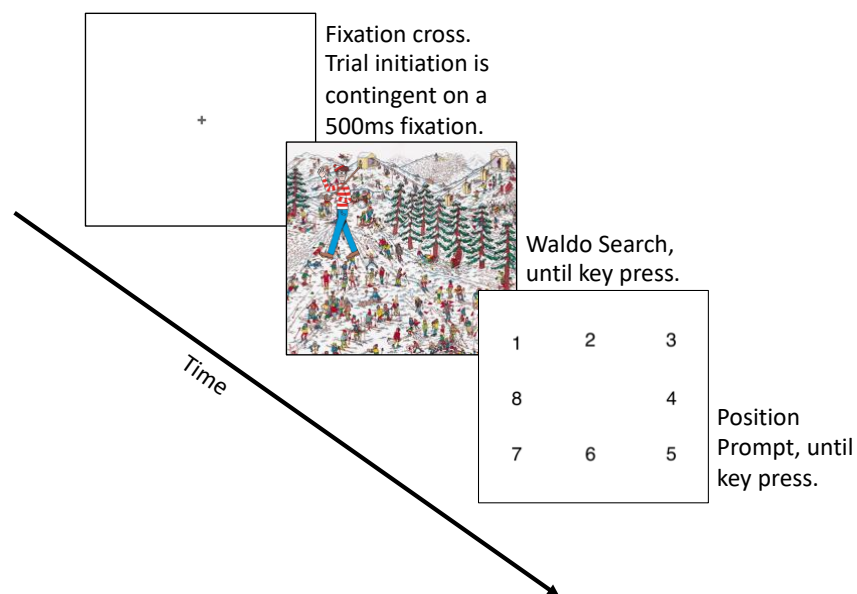
Procedure

Eye-tracking was performed using the web-based *Gazer* system. Participants were provided instructions to place the screen approximately parallel to their eyes and given visual feedback for head placement from a real-time video feed. Prior to the experimental trials, participants performed a 9-point self-calibration that was complete after they reached a validation threshold of 70% accuracy. Gaze predictions were recorded using the *Gazer* system for the duration of the experiment. The video feed of the laptop was continuously accessed to provide real-time predictions of screen-based gaze locations. Gaze predictions were calculated at a frequency of 30Hz.

Before each trial, participants were shown a gaze-contingent fixation cross. They proceeded to the Where's Waldo search scene once 500 *ms* of consecutive fixation was recorded on the cross. Immediately after, the background scene appeared (*i.e.*, trials started) with the Waldo target superimposed on top (Figure 2, slide 2). Participants were provided with 20 seconds to locate Waldo and press the [space] key to indicate he was found. The location of the target was one of eight possible locations (Figure 2, slide 3). After the [space] key was pressed, participants were shown a figure depicting the eight locations and asked to indicate the target's location with the corresponding key on their keyboard (*e.g.*, press the number [3] to indicate location 3).

Figure 2

An example of a trial timeline for the main Where's Waldo task.



Note. The Waldo target is enlarged in slide 2 for demonstration. The target's actual size was 152 *pixels* by 52 *pixels*. Participants fixated on a centered cross for 500 *ms*, then shown the Waldo background with the target presented randomly at one of 8 positions equidistant to center. They

would press [space] key once they located the target and then respond to a prompt asking for the Waldo's location in a 3 x 3 grid.

Each of the eight potential target locations appear once per block in random order. A new background was used for each block. Each session included a total of 8 backgrounds, for a total of 64 trials. Backgrounds repeated across sessions in random order. Participants completed the procedure in three separate sessions, between 24-48 hours apart.

Data Analysis

The current study leveraged multilevel modeling to evaluate how three dependent variables deviate across and within participants. The variables of interest included reaction time (RT, time to press the [space] key), number of fixations per trial, and fixation rate (number of fixations / second). Mean centering was performed on session. The three sessions were coded as session 0,1, and 2. The analysis was performed using the R module "lme4".

The fixed effects for slope in models for RT and fixations were modelled separately. These measures informed the initial analysis of performance. Individual differences for total fixations at baseline (*i.e.*, session 0) were assessed from the presence of random effects for the intercept of total fixations. Lastly, to probe fixations more closely, the random effects for fixation rate (fixations per second) were tested.

Results

Reaction Time

Based on an unconditioned model of RT (Model 1), participants performed the visual search task in an average of 7058 *ms*, *s.d.* = 326 *ms* [6731, 7384].

Reaction Time Across Sessions. The slope of and intercept of reaction time, conditioned on session, was calculated, and described in Table 1. Based on a likelihood ratio test, Model 2 (RT conditioned on session) was a better fit than Model 1 (unconditioned) ($X^2(3, N = 36) = 66.07, p < .001$).

Table 1

Slope and Intercept for Reaction Time to Press the [space] key, Conditioned on Session.

Effect	Measure
Fixed Effects (across participants)	
Intercept	8613** <i>ms</i>
Slope	-1932** <i>ms / session</i>
Random Effects (within participants)	
Intercept	2010 <i>ms</i>
Slope	170 <i>ms / session</i>

Note. The model indicates that participants completed the task in an average of 8613 ms in session zero, and performed 1932 ms faster in each subsequent session, for a total of 3864 ms reduced speed per trial by the final session.

** indicates $p < .05$. Random effects not tested for p-values.

RT was observed to decrease across sessions. Participants performed approximately 2 *seconds* faster in each subsequent session of Where's Waldo. Between the first (session 0) and final (session 2) sessions, participants improved their speed in the task by approximately 4 *seconds* per trial. Based on the average baseline of 8613 *ms*, this is an improvement of 55%.

Fixations

We evaluated participants eye movements as the total number of fixations made in each trial. Using an unconditioned model of total fixations (Model 1), the ICC (0.381) and design effects were calculated (1.607). The ICC determined interindividual dependency of measures. From the model, we observed variance was approximately 62% accounted for by within-persons effects.

Fixations at Baseline. Two MLM models were calculated to isolate the random effects for intercept. First, a model was created to estimate the random effects for intercept as zero, conditioned on session (Model 2a). A second model was calculated that estimated the random effects for intercept using maximum likelihood estimation (Model 2b). Models 2a and 2b were compared using a likelihood ratio test to directly test the significance of adding between-person variance (*i.e.*, random effects) to baseline performance. The output of the likelihood ratio test was significant ($\chi^2(2, N = 36) = 45.541, p < .001$), indicating participants' performance showed reliable individual differences at baseline.

Fixations Across Sessions. Using Model 2a, we estimated changes in total fixations per trial conditioned by session, or in other words, the changes in the number of fixations across the three testing sessions. The output of the fully conditioned model is shown in Table 2.

Table 2

Fixations Per Trial, Fully Conditioned (i.e., Intercept and Slope) on Session.

	Intercept	Slope
Fixed Effects	15.26 **	-3.56**
Random Effects	$4.77^2 = 22.76$	$0.71^2 = 0.50$

Note. The model indicates participants made an average of 15.26 fixations in the first session, and approximately 3.56 fewer fixations in each subsequent session.

** indicates $p < .05$ as calculated in model (random effects not tested)

The fixed effect for slope was estimated to be -3.56 ($p < .001$) fixations, therefore participants made reliably fewer fixations with each subsequent session. Compared to the first session, they made approximately 7 fewer fixations in each trial in the last session.

Fixation rate

To evaluate performance improvements more closely, an additional trial-level measure was created. Speed was calculated as: # fixations / (RT/1000). This value represents the average number of fixations made per second in one trial.

Fixation rate Across Sessions. Fixation rate was modelled separately to be unconditional (Model 1) and conditional on session for both slope and intercept (Model 2). Based on a likelihood ratio test, Model 1 (unconditioned) was a better fit than the conditioned Model 2 ($X^2(3, N = 36) = 1.69, p = 0.62$). To account for the worse fit, a third model was generated that was only partially conditioned (only slope) on session.

The partially conditioned Model 3 was a better fit than the unconditioned Model 1. Based on the output for Model 3 shown in Table 3, the fixed effects for slope were not significant. Therefore, participants did not show any speed improvements across sessions.

Table 3

Rate (Fixations Per Second) Partially Conditioned (Slope Only) on Session.

Effect	Measure
Fixed Effects (across participants)	
Intercept	1.797**
Slope	-0.001
Random Effects (within participants)	
Slope	0.213

Note. The model indicates participants did not show changes in eye movement rates across sessions, as the slope was non-significant for fixed effects.

** indicates $p < .05$. Random effects not tested for p -values.

Discussion

The goal of the current experiment was to record training effects in eye movements within a visual search paradigm. Participants performed a Where's Waldo task while their eye movements were being recorded using web-based Gazer program. Based on previous studies that evaluate visual search with training (Schuster *et al.*, 2013), we expected to find that participants would performance faster behaviourally at the key press response, and show greater rates of eye movements with subsequent sessions. To operationalize visual search performance, we

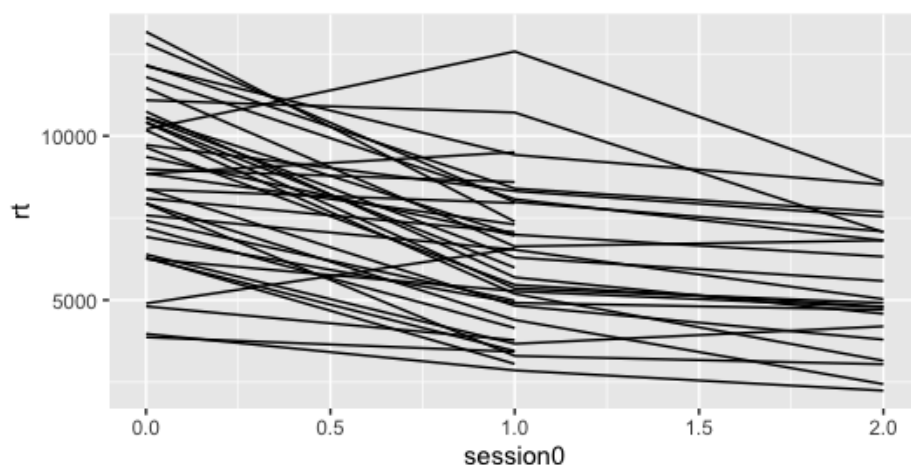
investigated changes in reaction time, total fixations per trial, and fixations per second (rate). We first investigated RT to establish general behavioural improvements.

Reaction Time

Participants demonstrated training effects for their reaction time to press the [space] key. On average, participant made the key response in each trial after approximately 8613 *ms* in session 0 and were 4 *s* faster by the final session. Reaction time across sessions is shown in Figure 3.

Figure 3

Average Reaction Time Per Trial Performed at Sessions 0,1, and 2



Note. Each line represents the trajectory of a single participant. On average, participants reduced their average reaction time per trial by 4 *s* between the first and final session.

This trend demonstrates training effects using our most straightforward measure of performance. However, it does not identify the source of the behavioural changes. Perhaps, the reduction in RT could be explained by a particular eye movement strategy. Specifically,

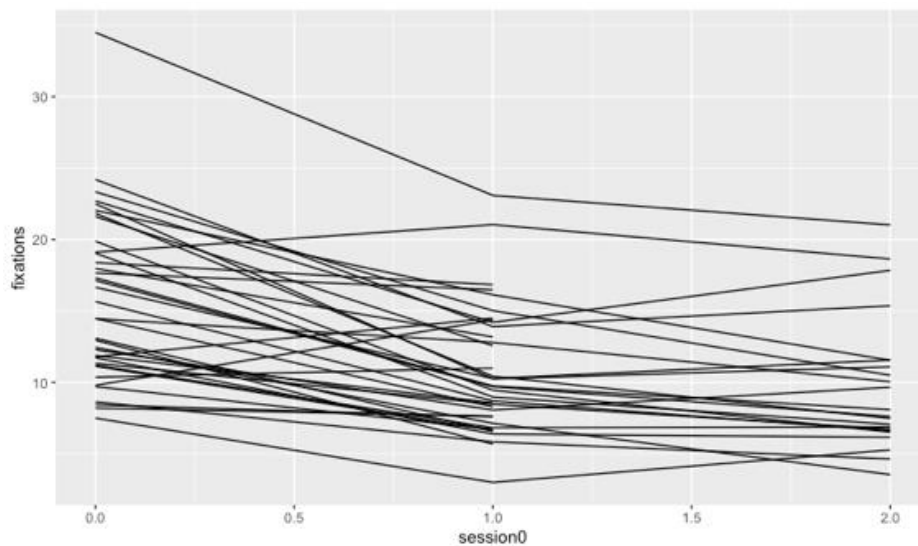
participants may be learning the eight potential locations of the target and orienting quickly to those locations, resulting in fewer fixations in each trial. To investigate this possibility, total fixations per trial were analyzed for comparison to the RT pattern.

Total Fixations Per Trial

The average number of fixations are shown in Figure 4 across sessions. A similar trend to the RT measure was present. Participants made between 3 to 4 fewer fixations with every subsequent training session.

Figure 4

Average Number of Fixations Per Trial Performed at Sessions 0, 1, and 2.



Note. Each line represents the trajectory of a single participant. Participants reduced the number of fixations by 3 to 4 with each training session.

Critically, this result provides a potential explanation for the behavioural improvement; participants begin to execute an eye movement strategy. During perceptual learning of the target

locations in session 0, they may develop a strategy to focus on the screen locations most likely to contain Waldo. No instructions were provided on how to perform the task, so adoption of a strategy was done spontaneously. As evidence of this strategy, participants made on average 8.14 fixations per trial in session 1, approximately equal to the number of target locations. In other words, the participants may have been fixating on each target location in session 2, until they found they found the correct area containing the correct Waldo target.

Individual Differences. To investigate the presence of individual differences in eye movements, we compared baseline performance for total fixations. In session 0, participants made approximately 15 fixations per trial. We recorded significant individual differences at this baseline ($p < .05$), as the addition of between-person variance increased the accuracy of our model for baseline fixations. These findings are in line with previous research showing that performance on a visual search task without training is quite variable (Joseph *et al.*, 2009). A closer inspection of individual differences indicated that the ranking of participants within in each session, based on number of fixations per trial, was stable across sessions ($r = 0.82$ ($p < .001$) between session 0 and 1, and $r = 0.89$ ($p < .001$) between session 1 and 2). Although participants showed a group trend towards fewer fixations across sessions, those who made the most fixations in the baseline session were also those who made the most in the final session. Perhaps, training does not affect relative performance in a visual search task.

Beyond looking at total fixations, fixation rate was calculated as measure to combine RT and total fixations to determine if eye movements were being made both faster and more efficiently.

Fixation Rate

If participant's fixation rate (fixations per second in one trial) increases, they may be showing search strategy changes beyond perceptual learning of Waldo and his locations. Total number of fixations were decreasing, but so was reaction time in this task, therefore the rate of eye movements may be consistent. Critically, this finding would differentiate between the case of participant's adopting the eye movement strategy of executing faster eye movements, versus choosing more strategic fixation locations. Fixation rate did not vary across sessions ($p > .05$). Therefore, our results indicate that the behavioural performance improvement may be a result of participants becoming more efficient at focusing their locations for fixations.

First Fixations

The locations of first fixations were probed across sessions and target locations to further investigate potential gaze strategies. Participant's first fixation locations were classified by the target Waldo location they fixated on. A main effect was present for location ($F(21) = 118.10, p < .001$), but not session ($F(21) = 1.60, p = .22$). Participants reliably looked at location 1 first, fixating in the top left location in approximately 33% of trials. However, this strategy was constant across all three sessions.

Search Task Strategy

In a further probe of strategy, participants may have partially executed a clockwise search, initiating search in the top left target location across all three sessions. In the final session, participants were also shown to make approximately 8 fixations per trial, the same number of possible target locations. We may speculate that by the final session, participants began searching in the upper left, then focused on each potential target location until they saw Waldo. This strategy may be augmented if participants maintain memory for locations of

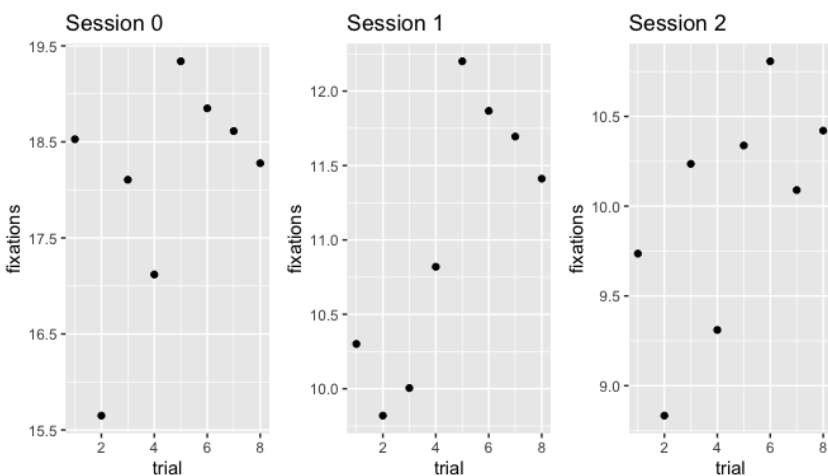
previous targets in the current block. The target locations each occur once per block, therefore when a target is found, the participants may infer that the target will not appear in that location again and there can eliminate it from search in subsequent trials. Indeed, when looking at fixations within a single block, participants appear to execute fewer fixations in the last few trials, indicating they are performing an elimination strategy (*i.e.*, ignoring locations they know no longer contain the target). An analysis of fixations indicated they were reliably different across trials. ($t(4960) = 3.60, p < .001$). In Figure 5 below, the average number of fixations in the trial sequence across blocks are shown for sessions 0 (left), 1 (middle), and 2 (right).

Interestingly, a pattern of reduced fixations across trials is potentially present in the final trials of each block. Particularly evident in sessions 0 and 1, participants improved after trial 5.

Participants may be holding the locations of previous targets in memory, but only act on the information after trial 5 when the array of remaining locations is smaller (and therefore easier to recall).

Figure 5

Average Number of Fixations Made in the Trial Sequence across Blocks in Sessions 0 (left), 1 (middle), and 2 (right).



Conclusions

We provide evidence that visual search performance can be improved through targeting eye movement efficiency. Participants were making fixations at a similar rate across sessions but were able to perform more quickly by being more intentional with their fixations. However, the relative performance of participants remained stable, indicating that those who are faster at baseline remain maintain their advantage, and repeated exposure to a visual search task does not eliminate this discrepancy.

In a further probe of strategy, participants appeared to execute a clockwise search, initiating search in the top left target location across all three sessions. By the final session, participants began searching in the upper left, then focused on each potential target location until they saw Waldo. When looking at fixations within a single block, participants appear to execute fewer fixations in the last trials in each block, indicating they are performing an elimination strategy (*i.e.*, ignoring locations they know no longer contain the target).

As a precursor to this search strategy, perceptual learning of the target locations must be performed. Our results indicate perceptual learning of a target-location combination can be done in only 40 visual search trials, and this effect compounds over subsequent training sessions. Interestingly, we were able to record individual differences in a visual search task and training effects across only 3-time sessions and using the web-based *Gazer* eye tracking program, confirming *Gazer's* utility in an attention task.

Future Studies. In future experiments, we will probe the mechanism behind the behavioural improvement further. We will add additional training sessions that will change the target stimulus locations. If performance reverts to baseline after moving the location (*i.e.*, fixation efficiency

does not transfer to new locations) then we may have evidence that participants are learning a Waldo-location combination as opposed to developing search heuristics for the Waldo target itself.

References

- Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2021). *MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture*.
<https://doi.org/10.31234/osf.io/rsdwg>
- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, 28(3), 241–248.
- de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of Experimental Psychology. Human Perception and Performance*, 14(2), 188–202.
- Drew, T., Evans, K., Vö, M. L.-H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics: A Review Publication of the Radiological Society of North America, Inc*, 33(1), 263–274.
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology. Human Perception and Performance*, 16(4), 802–811.
- Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., & Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder? *Developmental Science*, 12(6), 1083–1096.
- Schuster, D., Rivera, J., Sellers, B. C., Fiore, S. M., & Jentsch, F. (2013). Perceptual training for visual search. *Ergonomics*, 56(7), 1101–1115.
- vanWell, A. & Tanaka, J.W. 2023. "Gazer: A New Tool for Web-based Eye Tracking". In prep.

Williams, D. E., Reingold, E. M., Moscovitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, 51(2), 151–164.

Chapter 5: General Discussion

The goal of my work is to reflect on the viability of web-based eye tracking for psychological research. I have presented three separate studies that describe the *Gazer* system's capability to record separate components of attentional processing. Overall, our results suggest that *Gazer* is equipped to answer research questions related to both exogenous and endogenous attention. Critically, systems such as *Gazer* are enhancing research accessibility for scientists and enabling online data collection that is rapid and meaningful.

My work indicates that *Gazer* has the precision to discriminate eye movements between nine spatial regions on the screen. In Chapter 2, we present findings that *Gazer* is capable of recording saccadic eye movements from the center of the screen to one of the target spatial regions with comparable precision to the *Eyelink*. Previous studies (Simmelmann & Weigelt, 2018) have also shown that the underlying *Webgazer* (Papoutsaki, 2015) platform is capable of tracking moving objects (*i.e.*, pursuit eye movements). We expand on our findings in Chapter 3 to comment on endogenous eye movements. Critically, we find that region of interest (ROI) analysis can be performed on raw *Gazer* data to calculate measures including number of on-target fixations per trial, and first fixation onsets. In the final experiment (Chapter 4), we present findings that *Gazer* is capable of illuminating gaze strategy between participants and has fine enough precision to record at the level of individual differences.

Beyond research applications, systems such as *Gazer* may be suitable to answer non-academic questions. Individuals interested in how visual interest is captured on websites or by advertisements may be able to use this eye tracking technology to enhance marketing and sales techniques. *Gazer* may also be useful for participants as an alternative to the keyboard and mouse for computer interaction. Instead of selecting a button with a mouse, participants may be

able to interact with it through moving their eyes and focusing their gaze on the button's location. Indeed, recent advancements in technologies such as virtual reality (VR) headsets include this functionality, allowing the user to navigate the interface with solely their eye gaze. (Plopski *et al.*, 2022)

The primary goal of the *Gazer* project was to make the program easily available for researchers in the vision sciences community. By increasing the accessibility of web-based eye tracking technology, we can begin to see an even greater expansion of its applications and hone our understanding of the research questions that it is suited to answer. The program now exists as an open-source experimental package to be used in conjunction with the online experimental platform jsPsych. As with other open-source software, researchers have the capability to implement, modify and extend the *Gazer* code to meet the specific demands of their research program. The *Gazer* program is freely available for downloading at <https://github.com/amyvanwell> and includes jsPsych plugins and experimental templates.

References

- Papoutsaki, A. 2015. “Scalable Webcam Eye Tracking by Learning from User Interactions.” In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 219–22. CHI EA '15. New York, NY, USA: Association for Computing Machinery.
- Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., & Langlotz, T. (2022). The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *ACM Comput. Surv.*, 55(3), 1–39.
- Semmelmann, K, and Weigelt, S. 2018. “Online Webcam-Based Eye Tracking in Cognitive Science: A First Look.” *Behavior Research Methods* 50 (2): 451–65.