Online mouse tracking as a measure of attention in videos, using a mouse-contingent bi-resolution display

by

Karissa Payne

B.A., University of Findlay, 2019

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Lester Loschky

Copyright

Abstract

Data on human visual attention is increasingly collected online, but there are limited tools available to study attention to video stimuli in online experiments. Webcam-based eye tracking is improving, but it faces issues with precision and attrition that prevent its adoption by many researchers. Here I detail an alternative mouse-based paradigm that can be used to measure attention to videos online. This method uses a blurred display and a high-resolution window centered on the user's computer mouse location. As the user moves their mouse to view different screen content, their mouse movements are recorded, providing an approximation of eye movements and the attended screen location. To validate this Mouse-Contingent Bi-Resolution Display (MCBRD) paradigm, mouse movements collected from online participants watching twenty-seven videos were compared to eye movements from the DIEM dataset. Display settings of window size and blur level were manipulated to identify the settings that resulted in mouse movements most similar to eye movements. This validation study found differences in speed between mouse and eye movements, but similarities in attended regions of interest, especially when the MCBRD screen was blurred with the highest tested Gaussian blur sigma of 0.45 degrees of visual angle. These results suggest that the MCBRD paradigm can be used to measure what regions viewers find salient, interesting, or visually informative in online videos.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This work would not have been possible without the support and instruction I have received at Kansas State University. I would like to thank my Kansas State committee members Dr. Lester Loschky and Dr. Michael Young, alongside my outside committee member Dr. Rachel Brown from Nvidia, for all of the support, suggestions, and feedback they have provided. I would also like to thank my collaborators Dr. Brian Howatt (Kansas State University) and Sahand Shaghaghi (University of Waterloo) for their contributions in getting this paradigm up and running. This project was supported by the Cognitive and Neurobiological Approaches to Plasticity (CNAP) Center of Biomedical Research Excellence (COBRE) of the National Institutes of Health under grant number P20GM113109. Many thanks go to CNAP Director Dr. Kimberly Kirkpatrick for her support, along with Dr. Daniel Andresen, Adam Tygart, Dr. Dave Turner, and Kyle Hutson of the CNAP Neuroinformatics Core, who helped provide the High Performance Computing resources and related training necessary for this project.

Chapter 1 - Introduction

Over the last two decades attention researchers have begun studying viewers' attention while they watch videos (Carmi & Itti, 2006; Dorr et al., 2010; Mital et al., 2011). At the same time, people are spending increasing amounts of time watching video content on the Internet. The 2020 Nielsen Total Audience Report shows the average American spends at least 35 minutes a day watching video on smartphones, tablets, and computers alone, and over 4 hours watching video on tv or tv-connected devices (Nielsen 2020). This prevalence of video media raises the question: when people watch video content, where are they attending on the screen? The content we pay attention to is what we are generally aware of, can incorporate into our understanding, and later remember (Anderson & Pichert, 1978; Kaakinen et al., 2011; Madsen et al., 2012; O'Brien et al., 1988; Simons & Chabris, 1999). Knowing what people pay attention to when watching videos can inform us about their thoughts and cognitive processes. Online methods for collecting data on viewers' attention are increasingly sought after because data can be collected much faster than in-person data collection, data can be collected on a much larger scale, and data collection can continue even if pandemic conditions like COVID-19 prevent in-lab data collection.

How can we measure video viewers' attention while they watch videos online? The gold standard measure of the allocation of attention is eye tracking. However, high quality eye trackers are restricted to lab usage, and webcam-based eye tracking has not yet reached the point of having sufficient accuracy and precision to provide the quality of data needed by many attention researchers (Burton et al., 2014; Semmelmann & Weigelt, 2018). A simple alternative is the use of mouse-contingent multi-resolution displays (Anwyl-Irvine et al., 2021; Blackwell et al., 2000; Jansen et al., 2003; Jiang et al., 2015; Jones and Mewhort, 2004; N. W. Kim et al.,

2017). In such displays, a circular region of high-resolution is centered on the viewers' mouse location while the rest of the stimulus is presented in a lower resolution. To see something that is currently blurred, the viewer simply moves their mouse cursor to that location. In this way, the viewer's mouse movements provide an approximation of their eye movements, and thus, where they are attending on the screen. If implemented well, mouse-contingent multi-resolution displays should not cause any undue influence on the viewer's attentional selection compared to viewing the entire video image in high resolution. In this study I use prior research using eye tracking, gaze-contingent displays, and mouse-contingent displays to inform my current mouse-contingent multi-resolution display design. I then validate the optimal window radius and blur level settings necessary for obtaining data comparable to eye movements.

Current Measures of Attention

Eye tracking has been used as an empirical measure of attention since the late 1800s and has grown from the eye cap methods of Delebarre and Orschansky to the incredibly data-rich methods of today (see Wade, 2010 for a review of eye tracking history). Eye movements are a fruitful measure of attention because the eye *must* move (or saccade) in order to collect detailed information from different locations in the field of view. The distribution of rods and cones in the eye, ganglion cell convergence, and cortical magnification (more cells in the Lateral Geniculate Nucleus and V1 designated for processing central visual information) all result in our central vision having the highest resolution within our field of view, with diminishing resolution as distance from the fovea increases. Specifically, the higher density of cells in the fovea and the convergence of peripheral rods and cones onto fewer ganglion cells results in the highest perceptible spatial frequencies being in the fovea (1-2 degrees of visual angle eccentricity), followed by the parafovea (2-5 degrees of visual angle eccentricity) (Curcio & Allen, 1990;

Drasdo & Fowler, 1974; Quinn et al., 2018; Thibos, 1998; Watson, 2014). After making a saccade to a new location, the eye fixates to extract high resolution information. Our attention can be inferred from eye movements because attention cannot be disengaged from an active saccade. Each saccade to a new location is preceded by covert attention to that location (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995). Thus, both saccades and eye fixations can be recorded to inform us of where viewers have attended and may have extracted particular information from in a visual stimulus (Holmqvist & Andersson, 2017), the order in which they gathered this information (Bylinskii et al., 2018; Le Meur & Baccino, 2013; Noton & Stark 1971) and how long they processed this information (Findlay & Walker, 1999; Henderson, 2007; Just & Carpenter, 1976; Nuthmann, 2017; Nuthmann et al., 2010; Nuthmann & Henderson, 2012).

Nevertheless, it is not uncommon for circumstances to prevent the use of this gold standard measure of visual attention. Top of the line eye trackers are expensive (upwards of $10,000), and lower-level eye trackers ($1,000-$10,000) sacrifice spatial and temporal detail. Furthermore, data collection is limited to one participant per eye tracker at a time. The limitations of eye tracking data collection were made especially salient to many labs during the COVID-19 pandemic. With many labs shutting down in-person data collection, eye tracker use was not possible. Some of these labs shifted to online data collection in the form of webcam-based eye tracking and mouse-based methods of measuring attention (Anwyl-Irvine et al., 2021; Yang & Krajbich, 2021).

Webcam-based eye tracking (appearance-based gaze estimation) is improving as a methodology, but it is still not good enough for the detailed data collection required by many labs. Particularly, webcam-based eye tracking has calibration and precision issues (alongside

attrition issues that follow the difficulty of calibration-at-home) and is recommended only for studies that require rough estimates of eye movement locations, studies that are using large areas of interest, or studies with regions of interest that are not too close together (Burton et al., 2014; Semmelmann & Weigelt, 2018; Yang & Krajbich, 2021). Recent improvements in webcam-based eye tracking are due to the introduction of machine learning approaches, where models are trained on large datasets of images of various eyes, in different settings, looking at specific screen locations (George, 2019; Krafka et al., 2016; Zhang et al., 2015). These methods, although an improvement upon webcam-based eye tracking alone, still do not have the accuracy of commercial eye trackers and have not yet reached widespread use. The issue of low precision could be helped by collecting more data, similar to how a small statistical effect can be identified with a large enough sample size. However, the other limitations of webcam-based eye tracking (high attrition in particular) can make it difficult to gather a large enough sample to reduce the impact of variance from camera imprecision.

Another method of measuring attention in an online setting is through the use of mouse-based manipulation of the screen resolution without the use of a webcam. In particular, allowing a participant to see a region of high-resolution information only where they move their mouse has been shown to produce similar results to eye movements in terms of areas of interest visited and x-y coordinate similarity analyses (Anwyl-Irvine et al., 2021; Jiang et al., 2015; N. W. Kim et al., 2017). This methodology is extended and tested here in order to run online experiments with video stimuli and build upon previous gaze-contingent display paradigms. Reviewing the primary findings of tests of gaze-contingent display methodology can inform and guide this mouse-contingent paradigm.

Gaze-contingent Display History

The first application of gaze-contingent display methodology was in research on reading processes in 1975 (McConkie & Rayner, 1975). Generally, gaze-contingent displays change their display content based on the location of one's focus on the screen with a region of clear or high-resolution viewing most frequently tied to the location of the fovea, as identified by eye tracking (i.e., foveated rendering; Reingold et al., 2003). In the first gaze-contingent applications in reading research, text on a screen would change depending on where a participant was currently fixating. In this paradigm, legible text was only presented around the location of the readers' gaze (McConkie & Rayner, 1975). This method was used to test the extent of letters perceived around the reader's current point of fixation, but the process of changing screen content depending on gaze location would later be expanded upon in a variety of contexts.

Gaze-contingent displays have been implemented with letter scrambling in reading, blurred-periphery designs in scene viewing, and detail limitations in rendering 3D objects outside of the identified center of vision (Luebke et al., 2000). These latter two methods are largely directed towards saving image processing and transmission resources because high resolution content cannot be resolved by the visual system outside of the center of vision (Reingold et al., 2003). Thus, gaze-contingent displays have been used to save resources in many contexts, like driving and flight simulators, video teleconferencing, virtual reality, and augmented reality (Duchowski et al., 2004; Duchowski & Coitekin, 2007; Kim et al., 2019; Reingold et al., 2003).

The lower resolution content outside of the center of vision is designed to not be visibly different from a full high-resolution image to the viewer (i.e., a metamer in peripheral vision; Loschky et al., 2005). Gaze-contingent methodologies work because the fovea can resolve high

and medium and spatial frequencies whereas the periphery can only resolve low spatial frequencies (Cajar et al., 2016). If the resolution of a visual area is reduced, but still above the high-resolution cut-off (i.e., the highest resolution that can be perceived) at a given retinal eccentricity, the reduction will be imperceptible (Loschky et al., 2005) thus saving image processing or transmission resources without compromising the viewers' visual experience. The multi-resolution display is most commonly implemented by creating images at different resolutions beforehand or allowing an algorithm to filter the presented resolutions in real-time (Duchowski et al., 2005; Reingold et al., 2003).

These methods take advantage of the fact that the eye can resolve different spatial frequencies when they are presented at different retinal eccentricities. Psychophysical studies have identified functions describing the relationship (Peli et al., 1991; Pointer & Hess, 1989; Thibos, 1998; Watson, 2014). This function is captured by Geisler and Perry (1998) as:

$$CT(f\ ,e) = CT_0\ exp\ \left(\alpha f \frac{e+e_2}{e_2}\right) \quad (1)$$

where f is spatial frequency in cycles per degree, e is the retinal eccentricity in degrees, CT_0 is the minimum contrast threshold, α is the spatial frequency decay constant, and e_2 is the half-resolution eccentricity.

If the image degradation or blurring removes spatial frequency information that could have been resolved at any given eccentricity then that degradation will, in principle, be perceptible (Loschky et al., 2005). The information presented on the screen must adapt over time to the changing locations of the viewer's fovea. If updating the areas of clear and degraded visual information takes place during a saccade or even up to 50-70 ms after (Albert et al., 2017; Loschky & Wolverton, 2007), saccadic suppression prevents this update from being visible.

After the input methodology, the second largest difference between gaze-contingent displays and the use of mouse-contingent blurred displays is the purpose of the blurring. In the case of gaze-contingent displays, blur is designed to be imperceptible. For mouse-contingent displays, peripheral blur exists as a catalyst to spur movement of the high-resolution window. Why compare the two at all? The comparison to gaze-contingent displays is a crucial foundation, because the findings from research on gaze-contingent display methodology can greatly impact the design of a successful mouse-contingent display that captures attention similarly to how it would be allocated by the eye in an unblurred display. Specifically, the degree to which peripheral image content is blurred and the size of the unblurred foveal region can have a large impact on the attentional exploration patterns of the viewer.

Mouse-Contingent Multi-Resolution Displays

A mouse-contingent multi-resolution display is a viable method of obtaining data on participants' allocation of attention for one key reason: the blurred peripheral region of space prevents the viewer from gathering more detailed visual information unless they move the high-resolution window. In fact, visual blur is preattentively processed (Loschky et al., 2014; Peterson, 2016; Peterson, 2018). Clear information is more salient with uniquely clear content amongst blurred content capturing attention whereas uniquely blurred content will only capture attention if it is task-relevant (Peterson, 2018). Blur is also generally avoided by the visual system through lens accommodation. Lens accommodation (changing focus for objects at different distances) provides clarity to visual stimuli, as does making saccades to fixate on visual areas. These processes aim to reduce the discomfort that can be experienced with perceived blur (O'Hare & Hibbard 2013). In the mouse-contingent multi-resolution display, if a viewer wants to get detailed visual information about content on their screen, they must move their mouse to

reveal it. However, the peripheral content must be sufficiently blurred to motivate moving the window, but not so blurred that attentional selection is inhibited or discouraged. Blur that is too intense may prevent viewers from perceiving visual cues that could guide their mouse movements, similar to how we select our fixation locations from our low-resolution visual periphery.

Further motivation for thinking that a mouse-contingent multi-resolution display could serve as a measure of attention comes from the findings of studies that have investigated the correlation between eye and mouse movements in normal computer usage. Chen et al. (2001) recorded the eye movements of participants while they browsed websites and found a strong relationship between users' gaze and cursor positions (with eye gaze in the same region as the mouse at the beginning of a mouse movement 76% of the time, and at the end of a mouse movement 77% of the time). Other experiments have replicated this finding and have also shown that gaze and mouse cursor locations tend to be more correlated along the vertical (y) axis (Guo & Agichtein, 2010; Rodden et al., 2008). Cursor movements can differ from eye movements depending on the task if, for example, participants are reading text on a website versus clicking on screen content to navigate a page (Huang et al., 2012), but studies have shown that viewers commonly precede a cursor movement with a similar movement of the eye, or look at their cursor while moving it, while looking at full resolution screens. These results show both a strong connection between viewers' gaze and cursor locations, but also systematic temporal lags between gaze and mouse positions, with the gaze often leading the mouse to a new location.

Mouse-based multi-resolution displays have been used to measure attention since at least the early 2000s. They have since appeared in multiple forms, which differ in their method of viewpoint selection, online capabilities, and compatible stimuli.

Some multi-resolution display paradigms are based on mouse clicks on an image stimulus, revealing a window of high resolution on one clicked area at a time. These paradigms include Deng, Krause, and Fei Fei's online implementation of the BubbleView paradigm, where participants choose the location of clear-view bubbles, to aid in an object recognition task (Deng et al., 2013). The Bubbles paradigm, originally created by Gosselin and Schyns in 2001 with randomly placed bubbles of clear information on an otherwise obscured image, was adapted to be click-based by Deng et al. (2013) and allowed participants to choose the placement of the Bubbles. This expansion found results similar to free-viewing patterns in object recognition (Deng et al., 2013). This paradigm was expanded upon by Kim et al. (2017) in their BubbleView paradigm, which tested a click-based bubble paradigm on a wide variety of image stimuli (natural images, information visualizations, static webpages, and graphic designs). Kim et al. (2017) found that these clicks approximated eye fixations in defining areas of importance in stimuli. Kim et al. (2017) also found that their click-based paradigm resulted in a better approximation of eye fixations compared to SALICON (Jiang et al., 2015), a mouse-motion based paradigm. Click-based measures are certainly cleaner in terms of data collection—there is no need to parse continuous x-y coordinate data to identify mouse-saccades and mouse-fixations. Click-based methods would seem to be far less feasible for video stimuli, however. In their study only using static image stimuli, Kim et al. (2017) found slower exploration patterns in their collected mouse clicks vs eye movements. Clicks can potentially provide cleaner data, because they require conscious decisions on which area of the screen to reveal, but a window tied to mouse movements can potentially reveal different areas much faster, since there is no requirement for the viewer to consciously decide to click, and then do so. Instead, the viewer would only need to move their mouse to reveal the desired higher resolution. Note too that

mouse movement requires movement of the hand, which is assumedly a faster movement than moving a mouse and clicking with one's forefinger. This speed is especially important when considering the constant updating of video stimuli.

To date, most mouse-movement-based multi-resolution display paradigms have only been made for use with static image stimuli (for a comparison of mouse-movement methods, see Table 1). These methodologies date back to the early 2000s with the Restricted Focus Viewer of Blackwell et al. (2000) and Jansen et al. (2003), and a similar methodology used by Jones and Mewhort (2004). The Restricted Focus Viewer tied a rectangular window of high resolution to mouse movements, with four levels of increasing blur extending into the periphery. Blackwell et al. (2000) and Jansen et al. (2003) found exploration behavior similar to free viewing. Jiang et al.'s (2015) SALICON study used multi-resolution blur (increasing levels of blur into the periphery) on images tied to mouse movements, first in-lab and then extending online, and found resulting mouse-fixation maps to be similar to eye-fixation maps from the same images. When used online, SALICON required pre-study checks to ensure that participants' computers could run the multi-resolution blur design without lag. Multi-resolution designs that blur in real time are more computationally intensive than bi-resolution designs, which only require the processing of two resolution levels—the original resolution layer and the blurred layer.

MouseView.JS is a JavaScript library that allows for a bi-resolution display controlled by mouse movements in online experiments (Anwyl-Irvine et al., 2021). A Gaussian blur layer can be implemented on top of image stimuli, but video stimuli require the implementation of an opacity filter (overlaying a color of a set transparency). This opacity filter requirement is due to MouseView.JS working through a screenshot-and-blur method. The content of the screen is captured in a screenshot, a blurred version of this image is created, and then this blurred version

is drawn onto an HTML canvas overlay. This screenshot and blurred version is only updated when there is a user-initiated change in the underlying webpage by default (scrolling on a page, for example), but this updating can be set to instead update at a set time interval. The MouseView.JS team recommends an interval no faster than 1-2 seconds and instead using an opacity filter for video stimuli that update more frequently.

The experiment building platform PsychoPy has a mouse-contingent demonstration that can be used with video stimuli (https://gitlab.pavlovia.org/demos/dynamic_selective_inspect) employing an opacity-based filter. Opacity-based filters are less optimal than blur-based filters, as they can induce eye movements (mainly shorter saccades) that are different from normal viewing patterns. The online mouse-contingent platform FocalVid also uses an opacity-based filter for recording attention to video stimuli[1]. The only present literature describing an online, mouse-contingent multi-resolution display that allows for blur over video is a preprint from Lyudvichenko and Vatolin (2019), which describes mouse-contingent blurring of video working in a custom video player. However, this program is presently not available for open use, and a custom-video-player based implementation may limit the stimuli to only video and introduce complications when wanting to integrate this program with experiment creation software. The only presently available online methods that allow for video stimuli use opacity-based filters, which can differentially affect how attention is allocated in a scene. Specifically, reviews of relevant research using gaze-contingent methodologies find that more opaque filters can produce attentional exploratory behavior that differs considerably from the natural exploratory behavior of an unobstructed image.

[1] The next iteration of FocalVid, which I am a collaborator on, will add a multi-resolution blur method to their platform.

Table 1

Comparison of the Mouse-Movement-Based Multi-Resolution Display Paradigms Discussed

Tool	Stimulus	Resolution	Method	Available Online
MouseView.JS	Image	Bi-resolution	Blur	Yes
MouseView.JS	Video	Bi-resolution	Opacity	Yes
Restricted Focus Viewer	Image	Multi-resolution	Blur	No
SALICON	Image	Multi-resolution	Blur	No
PsychoPy Dynamic Selective Inspect	Image	Bi-resolution	Opacity	Yes
PsychoPy Dynamic Selective Inspect	Video	Bi-resolution	Opacity	Yes
Lyudvichenko et al., (2019)	Video	Multi-resolution	Blur	No
FocalVid	Video	Bi-resolution	Opacity	Yes
Current MCBRD Paradigm	Video	Bi-resolution	Blur	Yes

Attention in Multi-Resolution Displays

Knowing that mouse-contingent multi-resolution displays can serve as a measure of attentional allocation is good, but the best data would come from a display that most closely approximates natural viewing patterns. To approximate natural viewing patterns, the blur level and window size are likely to be crucial, and guidance may be provided by research on gaze-contingent methodologies.

Fine-tuning of the blur level in accordance with the function by which perceptible resolution drops off in the periphery is important to making the blur in gaze-contingent multi-resolution displays imperceptible (Reingold et al., 2003). However, in designing a mouse-contingent multi-resolution display paradigm to study attention, the fine-tuning of the blur level serves a different purpose: identifying the level of blur that induces participants to move their mouse but without causing them to search the screen area differently from the eye movement patterns they would produce when watching a normal unblurred video.

From research utilizing gaze-contingent display methodology, we know that saccade amplitudes decrease with greater peripheral blur (Cajar et al., 2016; Laubrock et al., 2013;

Loschky & McConkie, 2002; Reingold & Loschky, 2002a, 2002b; van Diepen & Wampers, 1998) or degradation in the form of spatial noise (Shioiri & Ikeda, 1989). In this case, the content in the high-resolution window is considered to be more salient and is more likely to become the target of the next saccade (in terms of the salience-based competition for saccade target location in the superior colliculus; see Findlay & Walker, 1999; Reingold & Loschky, 2002a, 2002b). The heightened salience of the content in the high-resolution window results in shorter saccades on average. Importantly, peripheral filtering similarly modulates covert attention (i.e., attention away from the fovea). Cajar et al. (2016) found that both central and peripheral scene degradation leads to inhibited target detection in the degraded region. Thus, excessive blur or degradation restricts both overt attention (i.e., eye movements) and covert attention.

Given the effects of excessive blur, it seems that the ideal blur level used in a mouse-contingent multi-resolution display should fall between two thresholds: the lower blur detection threshold, below which participants will not need to move their mouse, and the higher threshold, above which the blur begins to restrict both overt and covert attention (and subsequently, mouse movements). The blur detection threshold will depend on window size, with the just detectable blur threshold dropping to lower cycles per degree (higher blur) at larger eccentricities (Loschky et al., 2014). The threshold at which eye movements become unnatural is the point at which the level of blur outside the clear window is so high that eye movements become restricted to the clear window (similar to what was found in gaze-contingent displays). Because of the drop-off of visual resolution with increasing retinal eccentricity, this higher threshold of blur affecting attentional selection similarly depends on window size (Cajar et al., 2016; Loschky & McConkie, 2002; Nuthmann, 2014; Shioiri & Ikeda, 1998).

In validating the present mouse-contingent bi-resolution display paradigm, I tested three different levels of blur and three different window sizes in a 3 × 3 factorial within-subjects design to identify the combination of blur level and window radius that fit within these two thresholds and resulted in mouse movements that are the most similar to eye movements. I tested these blur and window settings in a variety of different video stimuli (pulled from the DIEM dataset, Mital et al., 2011), but it remains a possibility that stimulus differences may result in no clear best choice in terms of window and radius combination. Videos depicting scenes may be differently affected by blur that inhibits recognition of the gist of a scene (largely informed by the periphery, see Larson & Loschky, 2009). At the same time, too little blur (or too large of a window) could give viewers too much detailed visual information in some stimuli but not others, resulting in a decreased need to move the mouse. Blackwell et al. (2000) suggested using a level of blur that prevents participants from getting important information without moving a clear window of view over it (which can depend on both stimulus and task). In a similar paradigm, Jones and Mewhort (2004) suggested that a blur filter be customized to obscure diagnostic or biasing information. This blur level determination is similar to the method by which blur levels were decided in Kim et al. (2017). This validation will help to identify whether there is an optimal recommended setting for blur level and window size for video stimuli.

Hypotheses

This validation study comes with two sets of competing hypotheses regarding the similarity of mouse movements to eye movements and the generalizability of optimal blur and window settings across multiple video stimuli. Mouse movements are expected to be similar to eye movements, as was found in prior research (Anwyl-Irvine et al., 2021; Jiang et al., 2015; Kim et al., 2017), but the introduction of video stimuli could result in different exploration

patterns. Similarity between mouse and eye movements in terms of the distance explored, gaze similarity, and time spent within Areas of Interest (AOIs) will be analyzed to see if participants are attending to the same visual content on average, even if their exploration patterns differ. A lag between eye movements and mouse movements is anticipated due to the putatively inherent differences between moving one's eyes and moving a computer mouse. Eye movements are generally (though not always) automatic, require no conscious effort, and are quickly initiated. Conversely, moving a mouse is generally (though not always) a more controlled action which may require an intention to act, resulting in a delay in execution of movement (see a comparison of movement latencies for eyes and limbs in Abrams et al., 1990). There are also potential speed differences of eye movements versus cursor movements after the movements are initiated. This validation study tests for differences of mouse and eye movement speed through analyses of distance covered from each video frame to the next (see Distance Analyses below). The following analyses take potential differences of speed into account through temporal margins for error in the gaze similarity analyses (see Normalized Scanpath Saliency below) and the total time spent within AOIs (see Areas of Interest Analyses below).

It is hypothesized that the best blur and window settings could apply to most, if not all videos. This wide application of blur and window settings would be due to the combination of settings matching most naturally with the resolution of the visual system, and with the resulting blur falling between the blur detection threshold and the threshold at which eye movements are abnormally affected. As a competing hypothesis, it is also possible that blur and window settings are too stimulus-dependent for a pairing of settings to come out as a clear, optimal choice.

Chapter 2 - Methods

Materials: Mouse-Contingent Bi-Resolution Display (MCBRD) Paradigm

I implemented and tested a mouse-contingent paradigm based on mouse movement without clicking (in contrast to the BubbleView paradigm) because clicking may be slower and less natural for video stimuli. This paradigm is implemented in JavaScript (JS) and can be used in JavaScript-based experiments or alongside experiment creation software like PsychoPy that is able to use custom JavaScript code and export JS experiments to the web. This design utilizes the CSS (Cascading Style Sheets; a programming language for stylization of web-based content) blur backdrop filter that places a Gaussian blur with a designated standard deviation (blur sigma) in pixels (determining the strength of the blur) across the entire screen. A CSS clip-path is then set to the size of a circle indicated by the window radius value in pixels and is centered on a new location with each mouse movement. This clip-path allows us to cut through the blur layer and reveal the high-resolution content underneath. This method can be used over any type of visual stimuli, including video, and is relatively lag-free at 60 Hz (see appendix). The blur backdrop filter and clip-path are supported by most major web browsers, including Google Chrome, Mozilla Firefox, Microsoft Edge, and Apple's Safari browser. For this validation experiment, I will be using this code alongside an experiment created in PsychoPy (Peirce et al., 2019) on Pavlovia.org in an iFrame that can place this blur layer on top of any PsychoPy stimuli.

This paradigm is bi-resolution, meaning that there is a window of high resolution, and a uniform level of blur in the periphery. Two resolution levels (high resolution and lower resolution) are less resource-intensive than multiple different levels of blur which makes it better suited for running alongside video stimuli in an online experiment. A bi-resolution display (with the right window size) should not cause any decrement in similarity results compared to a multi-

resolution display. Kim et al. (2017) found that, when comparing a bi-resolution mouse-contingent paradigm to SALICON's multi-resolution blur, SALICON's results could be approximated without the multi-resolution blur thus saving computational resources. Loschky and McConkie (2002) tested a gaze-contingent bi-resolution display with a high-resolution window tied to the participants' gaze and a lower resolution area in the periphery. They found a larger number of shorter saccades when there was heavy peripheral filtering, but they did not find that saccades were abnormally allocated to the window edge eccentricity. Furthermore, Loschky and McConkie (2000) found that bi-resolution displays with a sharp edge did not increase peripheral degradation detection compared to a blended-edge design. In related research, Hoffman et al. (2018) found hard transitions between resolutions in the periphery to be less visible than smoothed transitions in VR foveated rendering. Finally, Reingold and Loschky (2002) found that peripheral targets were located more slowly when the periphery was degraded, but this finding was unaffected by whether there was a blurred versus sharp window edge. These results suggest that the blurred screen content, set to a blur level intended to keep eye and mouse movements relatively natural, should not be adversely affected by the edge of the mouse-contingent window.

To address internet lags that can affect video-based experiments online, videos were loaded into the browser cache before participants could begin the experiment. Despite introducing a wait time before the beginning of the experiment (3-5 minutes in testing), pre-loading videos into the cache lowers the chances of video lag during experimentation. Likewise, it was recommended that participants close all other programs and browser tabs before beginning the experiment to reduce any lag that could be caused by the computer's RAM and CPU resource limitations.

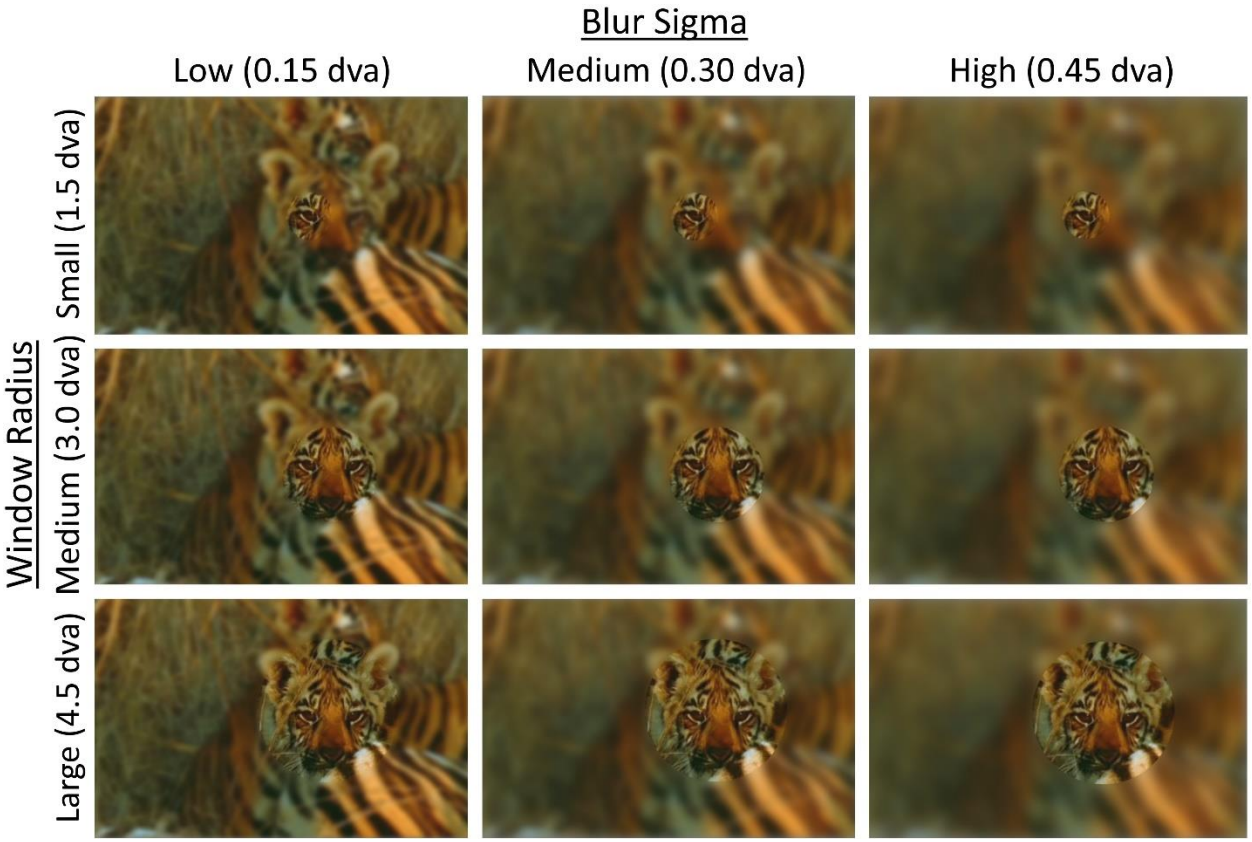
Materials: Stimuli

Participants watched 29 1- to 3-minute videos, all presented through a mouse-contingent bi-resolution display, with different combinations of three blur levels and three window sizes in a 3 × 3 factorial within-subjects design. All participants experienced each of these nine combinations three times. The first two videos were used as practice to reduce the variability that could come from participants getting acclimated to the mouse-contingent display. These practice videos were not part of the 3 × 3 design and were run with middle values of both blur sigma (0.3°) and window radius (3.0°). To test window sizes close to the sizes of the fovea, the middle of the parafovea, and the beginning of the visual periphery in the 3 × 3 design, I used window sizes of 1.5°, 3.0°, and 4.5° of visual angle, respectively. These window radii are close to the window sizes of 1.6°, 2.9°, and 4.1° used by both Loschky and McConkie (2002) and Nuthmann (2014) in gaze-contingent window designs with a slightly larger range to strengthen the difference between the window size options. These window sizes will test the effect of allowing different amounts of the parafovea to be presented without blur. For window size visualization, see Figure 1.

The blur levels are characterized by the standard deviation of the Gaussian blur applied (blur sigmas). Here I have implemented blur sigmas of 0.15°, 0.30°, and 0.45° of visual angle. These blur sigmas result in visibly similar blur to two of the wavelet-created blur conditions used in Loschky and McConkie (2002) with an additional stronger blur condition to test the similarity of movements induced by a stronger blur. For blur level visualization, see Figure 1.

Figure 1

Visualization of All Window Radius and Blur Sigma Level Combinations



The visual angle values of the window radii and blur levels were converted to pixels depending on each individual participant's screen size and distance from the screen (using an adaptive screen scaling method described below). Existing mouse contingent display paradigms have not used visual angle in determining their window radius and blur sigma levels—these paradigms are all solely pixel based, or based on a proportion of screen size, which does not take viewer distance into account. Basing display content on pixel number or screen proportions is problematic because the resulting physical stimulus size will likely be different for each participant. BubbleView provides the visual angle equivalents of their pixel values, but these visual angle values are based on the eye tracking condition their comparison data was collected

in—BubbleView itself was tested using downsized videos with Internet participants using unmeasured screen sizes and watching from unmeasured distances to the screen (Kim et al., 2017). This paradigm may highlight the utility of using visual angle in setting display sizes for online participants in different viewing conditions.

After watching two practice videos, participants viewed all 27 test videos and experienced each of the nine blur level and window size combinations three times. The conditions each participant experienced were randomly selected from a data source with conditions block randomized such that each video was experienced in each viewing condition an equal number of times by the end of the study. With this counterbalancing, each video was experienced with different window sizes and blur levels by different participants. The videos, along with the existing eye tracking data from 42 to 220 participants per video, are from the Dynamic Images and Eye Movements (DIEM) dataset (Mital et al., 2011), a dataset popular for its use as ground truth in testing saliency algorithms for videos (Dorr et al., 2012; Gorji & Clark, 2018; Tangemann et al., 2020; Wang et al., 2021). All available DIEM eye tracking data for each video was retained to provide the most informed baseline of eye movements. Unequal DIEM sample sizes were accounted for in creating the baselines in our analyses such that videos with more participant data did not influence the baseline means more than videos with less participant data (see Results below).

Mital et al. (2011) originally used 26 videos, so I included one more video from the DIEM dataset to reach 27 total test videos (to fit the 3 × 3 design). Two other videos were included as practice. These videos included a wide range of video stimulus types, including documentary footage, advertisements, trailers, and time lapse footage. All videos were presented at 30 frames per second, at a resolution equal in visual angle to the original video pixel

resolutions (ranging from 704 to 1280 pixels in width, and from 480 to 720 pixels in height) presented in the original in-lab viewing conditions (21'' monitor with desktop resolution of 1280 × 960, and a viewing distance of 90cm). Videos were compressed to 25% of their original bitrates—a percentage of compression found for these videos that lowers video file size (and subsequently pre-experiment load time) without noticeably hindering the visual video quality. The details of all videos used are available in Table 2.

Table 2

Video Stimuli Used From the DIEM Dataset

Video	Video type	Width (px)	Height (px)	Width (dva)	Height (dva)	Duration (s)	Frames
1	Advertisement	1024	576	21.3	12.1	41	1216
2	Advertisement	1024	576	21.3	12.1	40	1200
3	Advertisement	1280	720	26.7	15.1	72	2164
4	Advertisement	1272	720	26.5	15.1	30	898
5	Documentary	1280	720	26.7	15.1	109	3281
6	Documentary	1280	720	26.7	15.1	99	2968
7	Documentary	1280	720	26.7	15.1	145	4359
8	Documentary	1280	720	26.7	15.1	106	3180
9	Documentary	1280	720	26.7	15.1	87	2602
10	Documentary	1280	704	26.7	14.8	169	5080
11	Game trailer	1280	720	26.7	15.1	124	3718
12	Game trailer	1280	720	26.7	15.1	103	3100
13	Game trailer	1280	720	26.7	15.1	110	3313
14	Game trailer	1280	548	26.7	11.5	181	5418
15	Home movie	960	720	20.0	15.1	55	1660
16	Movie trailer	1280	690	26.7	14.5	109	3282
17	Movie trailer	1280	688	26.7	14.5	100	2992
18	Music video	880	720	18.3	15.1	30	886
19	Music video	1024	576	21.3	12.1	187	5595
20	Music video	1280	720	26.7	15.1	43	1280
21	News	768	576	16.0	12.1	102	3071
22	News	768	576	16.0	12.1	67	1998
23	News	1080	600	22.5	12.6	86	2571
24	News	960	720	20.0	15.1	209	6275
25	News	768	576	16.0	12.1	166	4978
26	Time-lapse	1280	720	26.7	15.1	47	1417
27	Home movie	960	720	20.0	15.1	39	1169
28	Documentary	1276	720	26.6	15.1	144	4319
29	News	1280	720	26.7	15.1	80	2409

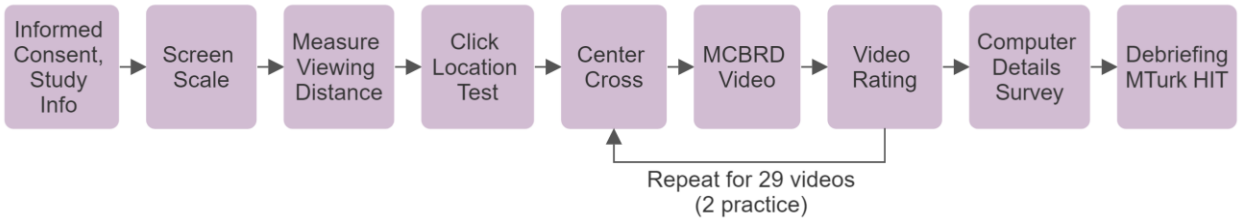
Note. Video size is reported in pixels and degrees of visual angle (dva) according to the screen size and distance reported in Mital et al., 2011. Videos 28 and 29 were used as practice videos. The names of all videos as they are listed in the DIEM dataset are featured in Appendix E.

Procedure

A total of 54 participants were recruited through Amazon's Mechanical Turk platform using Cloud Research's MTurk Toolkit (for better participant pre-screening options to allow for higher quality data collection; see Eyal, et al., 2021). I chose this sample size to be evenly divisible by nine, so that each blur and window size ordering in the 3 × 3 factorial design was experienced by six participants per video. Similar within-subjects designs are adequately powered by 50-60 participants. Each participant was compensated $15 for the experiment which lasted roughly an hour. This study had multiple steps as illustrated in Figure 2. Each of these steps is described in detail below.

Figure 2

Procedural Steps Encountered by All Participants in This Study



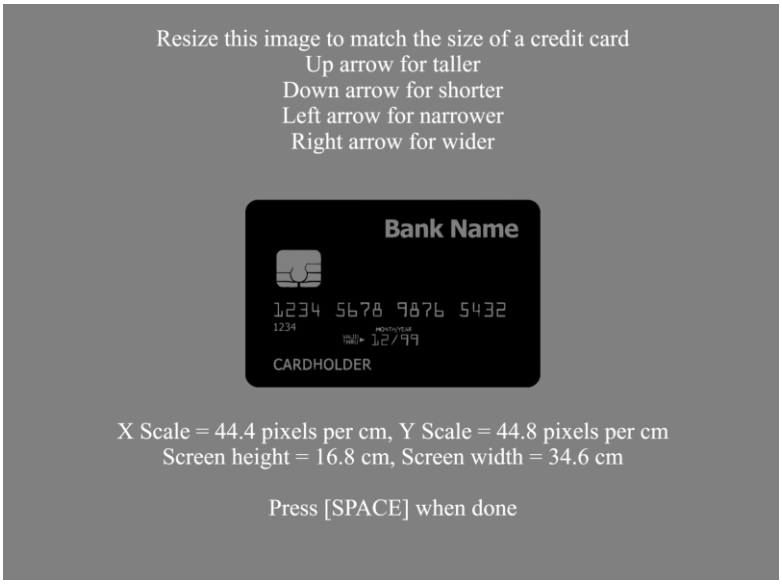
Participants were first given a link to a Qualtrics page that included the informed consent for the study, a description of the study, demographic questions (age and sex), and a place to report their MTurk ID. Participants were then asked if they had a card the size of a credit card and a tape measure to use in the screen size estimation (explained in more detail below), and measuring their viewing distance. Once participants verified they had both items ready, they were given a link to the PsychoPy experiment on Pavlovia and informed that they should not exit

the fullscreen window once the experiment began (screen size throughout the experiment was collected and reviewed to ensure participants adhered to these instructions—see Results below.).

After all videos had loaded into the browser cache and participants had entered their MTurk ID on the experiment page, the system estimated the number of pixels per degree of visual angle for the participant's viewing setup. Using Morys-Carter's ScreenScale code on Pavlovia (2021), participants were asked to place a credit card (or card of similar size) against their screen and rescale an on-screen rectangle until it matched the size of the physical card (see Figure 3). The number of pixels equal to one centimeter was calculated for the participant's display.

Figure 3

Card-Based Screen Scale Method

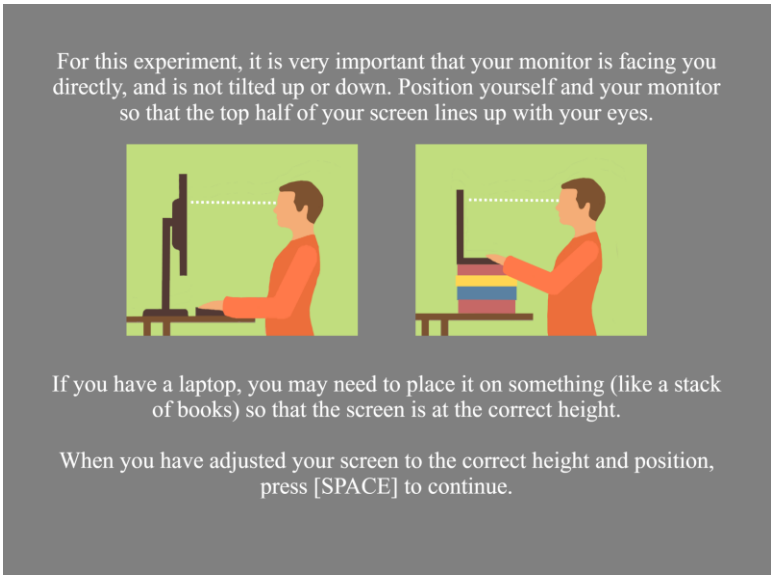


Calculating the visual angle of content on the screen also required knowing the participants' distance from the screen. Participants were given instructions regarding screen

placement to ensure that their screen was facing them directly and that the top half of their screen lined up with their eyes. The “top half” was given as an instruction for participants so the screen height would be comfortable for laptop users who would be propping up their devices such that they would not need to keep their arms raised as high to use their trackpads. This “top half” instruction also ensured that participants would still be facing the screen directly, even if they eventually slouched over the course of the hour-long study. See Figure 4 for the instructions as they were presented to participants.

Figure 4

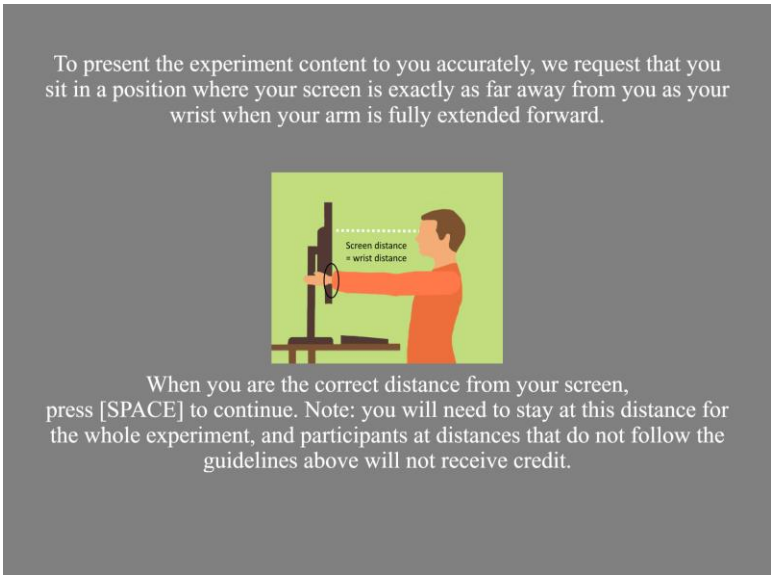
Screen Height Instructions That Were Presented to Participants



Participants were then instructed to sit at an arm’s length from the screen. An arm’s length is a comfortable viewing distance that did not result in participants sitting too close to their computers and that was easy to describe without instructing participants to match their viewing distance to exact inch or centimeter values. These instructions are shown in Figure 5.

Figure 5

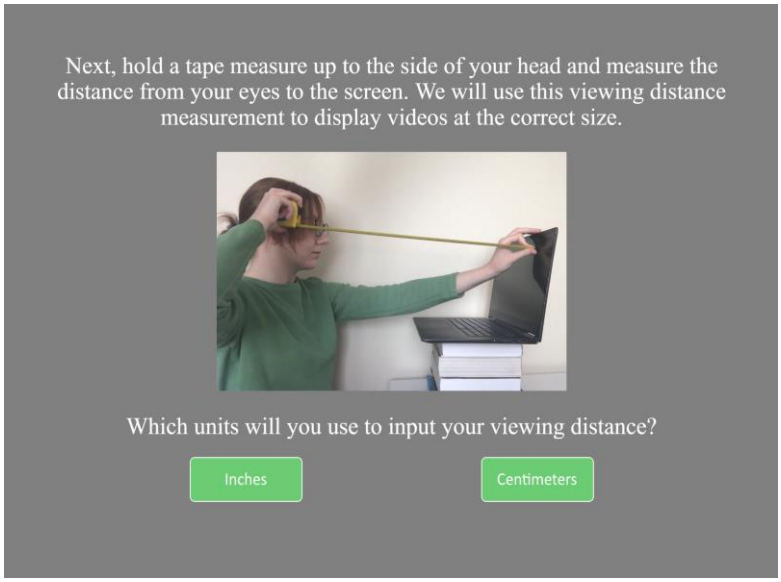
Screen Distance Instructions That Were Presented to Participants



After participants established their viewing height and distance, they were instructed to use a tape measure to measure the distance from their eyes to the screen. Participants were presented with an instructional video detailing how to correctly make this measurement. A still of this video is presented in Figure 6. Participants were asked to input their viewing distance in inches or centimeters, and they were then presented with a screen asking them to confirm that their measurement was correct. A broad interpretation of arm length range was used to determine if participants followed instructions and sat an arm's length away from the screen. Participants were removed for reporting distances shorter than 38 cm (15 inches), and distances longer than 101 cm (40 inches).

Figure 6

Screen Distance Measurement Instructions That Were Presented to Participants



Participants' viewing distances were converted to centimeters (if they were not already entered in centimeters) and then used with their unique pixel-to-centimeter ratio (calculated in the card scaling task) to calculate the pixels necessary to display content at specific visual angle sizes throughout the experiment. For this calculation I used the following formula:

$$Pixel\ output\ = tan\left(\left(\frac{\pi}{180}\right) \times \left(\frac{VA}{2}\right)\right) \times 2 \times d_{cm} \times w_{screen\ (px)} \,/\, w_{screen\ (cm)} \qquad (2)$$

where *VA* is the desired visual angle, *d* is the measured distance from the screen, and *w* is equal to the width of the screen (in pixels and centimeters in the ratio above). All videos were presented at the same size in degrees of visual angle as they were originally presented in the DIEM study (Mital et al., 2011; see Table 2).

After the participants' pixel-to-centimeter ratio and distance from the screen were collected, participants were asked if they were using a computer trackpad or mouse. Next, participants were presented with a mock eye-tracking calibration screen with instructions to click

on nine fixation crosses as they appeared in a mouse contingent blurred display (the "click location test" in Figure 2). This mock calibration screen served as the first introduction to the mouse-contingent display method, and the clicked locations were used as a check to verify active participation (participants would be removed if they did not follow instructions and move their mouse to different cued regions of the screen). The screen advanced when the mouse was clicked but only clicks reasonably near the center of the fixation crosses (within the 45 × 45 px area occupied by the cross shape when coordinates were transformed back to the 1024 × 768 size of the original image), signified an attentive participant whose display truly matched the pixel size identified by PsychoJS (PsychoPy's JavaScript library) *win.size* function. This step ensured that the dimension of the blur window matched the dimensions recorded by the screen.

Participants watched two practice videos (videos 28 and 29 in Table 2) to familiarize themselves with the mouse-contingent display process. After completing two practice videos, participants watched the 27 trial videos in a randomized blur level and window size condition (with no condition being used more than three times per participant). Each video was preceded by a clickable cross shape in the center of the screen that functioned the same as the fixation dot preceding the videos in the DIEM eye movement data—this clickable cross regularized the starting location between participants. After each video, the participant was asked how much they liked the video on a scale of 1 (did not like) to 4 (liked a lot), similar to the procedure in Mital et al. (2011), to add interactivity and additional motivation for attending to the videos.

After participants finished watching all videos, they were directed to a Qualtrics survey where they were asked questions about their computer usage and computer setup. These questions were used to describe the nature of the sample. Participants were asked how much experience they have with computer-based video games on a scale of 0 (no experience) to 5 (a

lot of experience), whether they were using a laptop or a desktop, whether their computer was connected to multiple monitors, their operating system, and their browser. Participants were asked if they were wearing bifocals, and they were asked for their height in feet and inches (used to estimate what a reasonable range was for the arm's length viewing distances). Participants were also directed to find their mouse speed and screen refresh rate in their system settings and report these values. Finally, participants were asked if they encountered any issues, if they were confused by anything in the study, and if they had any feedback to provide about their experience. They were then given a code to redeem their $15 MTurk payment.

Refresh Rate Measurement

After the conclusion of the study, I created a second experimental program to test the refresh rate of the MCBRD display. Using the existing MCBRD paradigm, I created a program with one looping video presented with a mouse-contingent window. The visual angle calculation procedures were skipped for this testing program, and instead blur sigmas and window radii of 50 px each were implemented for simplicity when testing refresh rates in different browsers and on different computers (this procedure required frequently exiting and restarting the program in different settings). Although participants in this study experienced blur intensities and window sizes that depended on their screen's size in degrees of visual angle, these 50 px testing values appeared to be a high amount of blur and a small window size. These values were chosen to test a higher amount of blur, in case it became more computationally intensive, to ensure the refresh rate measurements were conservative. A 500 × 500 px black square was overlaid on top of a 500 × 500 px white square in the top left corner of the screen. With each detected mouse movement, the transparency of the black square would shift from 1 (fully opaque) to 0 (fully translucent, with the white square then showing underneath). A still from this program is featured in

Appendix C. I placed a photodiode (interfacing with a Chronos response box) on this square and affixed it with elastic. When I moved the mouse continuously, the box would flicker between black and white. This flicker rate would thus be equal to the rate at which the blurred screen presentation updated. I measured the refresh rate for three different computers, and in three browsers per computer. Different mouse types and screen resolutions were also tested on one computer. The details of these refresh rate measurements are available in the Appendix.

Chapter 3 - Results

Sample Demographics

This study originally recruited 66 participants, however four were excluded from the analyses because they quit the experiment after the calibration routine (not providing any mouse movement data) and eight more were excluded for entering implausible viewing distances (see exclusion criterion in the Methods section). No participants were removed for exiting out of the full screen experiment or for not following instructions in the click location test. Subjects were recruited until a final sample size of 54 was reached (using all datasets in the block-randomized 3 × 3 factorial design).

Of the participants that completed the study and had input distance measures that plausibly fit the instruction of sitting roughly one arm's length away from the screen, some participants had their data for select videos removed for potential lag issues. I defined unacceptable lag as total time spent in the video PsychoPy routine lasting longer than 2 seconds past the duration of the video (allowing the potential for brief data transfer time at the end). Participants with potential lagged videos had their data for the lagged videos removed in the analyses. One participant had their data for three videos removed. Two participants had their data for two videos removed. Two participants had their data for one video removed.

Within the final sample, 21 participants were female and 33 were male with a mean age of 41 (SD = 12). The computer-related information for the 54 final participants is included in Table 3. For the responses on a continuous scale, means and standard deviations are included in Table 4. There are two things to note in these results. First, other than a similar number of participants using laptops and desktops, the responses were highly skewed towards a common setup—most participants used the Chrome browser on a Windows computer, used a computer

mouse opposed to a trackpad, and did not wear bifocals. There were also some unexpected options reported (like the Opera browser, or a participant reporting using a Trackball). The data from these participants did not appear different from the others and were kept in the final analyses. Future work can use these variables as predictors to determine any potential impact of different technological conditions, but sample sizes were too small for such an analysis in this study. Future studies could experimentally manipulate these conditions (other than bifocal usage) to directly examine the impact of these factors on the MCBRD output.

Table 3

Categorical Computer and User Information

Measurement		*N*
Computer Type		
	Desktop	28
	Laptop	26
OS		
	Windows	47
	Mac OS	4
	Linux	3
Browser		
	Chrome	44
	Edge	3
	Firefox	6
	Opera	1
Bifocals		
	No	51
	Yes	3
Mouse Type		
	Trackpad	7
	Mouse	46
	Trackball	1

Table 4

Descriptive Statistics for Continuous Computer and User Information

Measurement	*M*	*SD*	Range
Screen Refresh Rate (Hz)	66.5	22.7	40–144
Mouse Cursor Speed Percent (Windows)	51.5	14.7	5–100
Mouse Cursor Speed Percent (Mac)	52.5	25.0	40–90
Gaming Experience (0 to 5 scale)	3.4	1.4	1–5
Screen Width (pixels)	1755.2	500.9	1080–3440
Screen Height (pixels)	985.9	210.9	576–1728
Screen Width (cm)	43.9	12.1	25.3–92.0
Screen Height (cm)	24.9	7.0	14.4–51.1
Distance from screen (cm)	56.7	9.2	38.1–78.7

Data Pre-Processing

The initial mouse cursor position results in x-y coordinates are in terms of the participant's computer screen size. Additionally, the mouse position is recorded at 60 Hz, but the position is only recorded/updated when the mouse moves. Mouse positions when the mouse does not move therefore require interpolation. Initial interpolation extends the data to a sample every millisecond—this interpolation can then be downsampled to 30 Hz to match the downsampling used in the DIEM dataset, and the frame rate of the DIEM videos. The raw x-y coordinates of the 30 Hz samples were then transformed in terms of each participants' unique video size using the following equations where w is width in pixels, and h is height in pixels (for both the participant's screen and the video size, see below):

$$x_{transformed} = x - \left(\frac{w_{screen}}{2}\right) + \left(\frac{w_{video}}{2}\right) \quad (3)$$

$$y_{transformed} = y - \left(\frac{h_{screen}}{2}\right) + \left(\frac{h_{video}}{2}\right) \quad (4)$$

After participants' x-y position data are transformed into coordinates respective to their individual video sizes, these coordinates are taken as a proportion of the participants' video size in width and height respectively:

$$\left(\frac{x_{transformed}}{w_{video}}\right), \left(\frac{y_{transformed}}{h_{video}}\right) \quad (5)$$

These values are then multiplied by the original display size of the video in pixels:

$$\left(\left(\frac{x_{transformed}}{w_{video}}\right) * w_{original}\right), \left(\left(\frac{y_{transformed}}{h_{video}}\right) * h_{original}\right) \quad (6)$$

These normalized x-y values allow the data to be compared among all mouse-contingent bi-resolution display (MCBRD) participants and compared with the DIEM eye tracking data for each video.

DIEM Baselines

For the following analyses, direct comparisons were only made between MCBRD conditions in the 3 × 3 design. The eye tracking data, collected from an entirely different sample, only serves as a baseline value that can be used to interpret the MCBRD results. DIEM mean baselines were gathered by running null models for each analysis on the DIEM data with no fixed effects and only random effects of video intercept and participant intercept. This method allowed for means (e.g., distance traveled per frame) to be estimated while controlling for different video lengths and different DIEM sample sizes per video.

The following results will be interpreted in two ways. First, if a MCBRD condition is deemed to create behavior significantly closer to the DIEM baseline than other conditions in these analyses, that condition will be deemed to generate the closest results possible to eye movements. However, this significance is qualified by the distance to the DIEM baseline. If a setting condition results in significantly closer values to the DIEM baseline than other MCBRD

conditions, but the DIEM baseline is still farther away from all MCBRD values, the significant difference between MCBRD conditions may not result in an important difference. These results are presented such that researchers can use the settings that best fit their goals, and the sizes of these differences may have different interpretations for different researchers. Here I suggest the following: If a difference between MCBRD conditions is smaller than the difference between the lower and upper 95% confidence intervals for the DIEM baseline, the difference, though potentially significant, may not be meaningful (the difference would be similar to the natural differences seen in participants viewing videos with an eye tracker). Furthermore, for the Area of Interest analyses (detailed below), differences between MCBRD conditions are meaningful if the MCBRD AOI results on average are closer to the DIEM baseline than they are to chance (the average likelihood of viewing an AOI on any given frame). For the Normalized Scanpath Saliency analyses, if a MCBRD setting results in an NSS z-score greater than zero on average, that setting has gaze similarity that is greater than average to the DIEM ground truth. Differences between MCBRD settings with results above this zero value are considered to be meaningful. For the distance per frame analysis, this interpretation is more subjective and depends on the intended use case of the MCBRD paradigm. Here, I suggest that values closer to the DIEM baseline than half of the baseline are meaningful to interpret. Whether the following results meet these criteria will be discussed for each analysis.

Finally, the goal is to find settings that result in mouse movements closest to eye movements. This mouse-contingent display paradigm can of course serve as a unique measure of attentional selection interpreted aside from eye movements. If a researcher prefers to use this paradigm without consideration of the comparison to eye movements, only the differences

between the 3 × 3 MCBRD settings in the Distance and AOI conditions are meaningful (the NSS analyses are based on comparisons to the DIEM ground truth).

Distance per Frame Analyses

Gamma Regression

Distance traveled between subsequent x-y position samples was calculated in Euclidean distance:

$$Distance\ traveled = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (7)$$

After I calculated the Euclidean distance between each sample, I converted this distance back to visual angle using the pixels-to-visual-angle ratio obtained from the original DIEM data video width in pixels and visual angle:

$$Euclidean\ distance_{dva} = \left(Euclidean\ distance_{px} * \left(\frac{w_{dva}}{w_{px}}\right)\right) \quad (8)$$

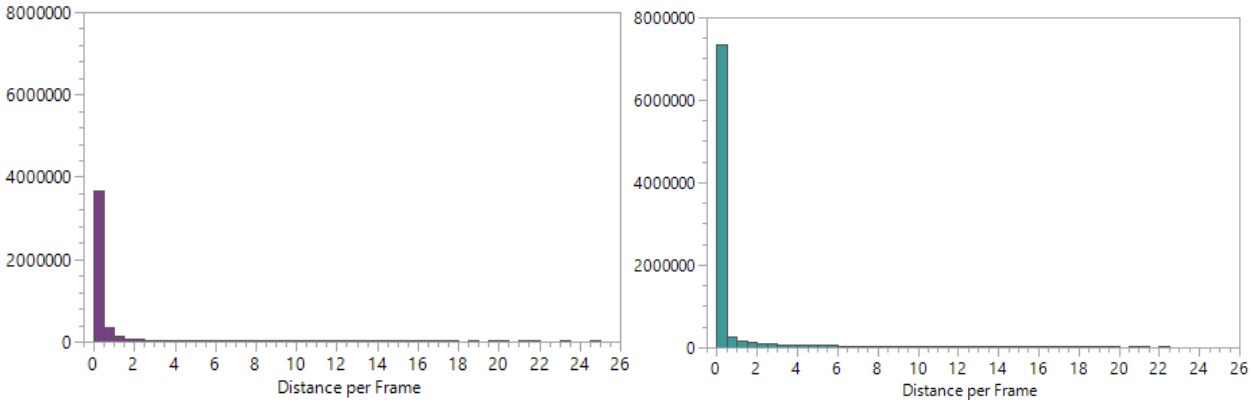
Although measures of distance, velocity, and acceleration (the variance of which is addressed in these analyses through the measure of distance covered over each 30 Hz video frame; velocity and acceleration are derivations of this measure) are most frequently used in identifying saccade, fixation, and smooth pursuit events (Holmqvist & Andersson, 2017), the threshold values for determining these events are only established for eye tracking data. Future work will determine what these thresholds may be for mouse tracking data, and the process of determining thresholds will be elaborated upon in the Discussion. For the analyses of this study, I will only be comparing distance per frame as values with no defined event titles.

I conducted all analyses in R (version 4.2.1) using the lme4 library (Bates et al., 2015), and using the emmeans library (Lenth, 2019) to calculate the estimated marginal means for each model and to conduct contrast tests. The Euclidean distance data showed a heavy right skew (see Figures 7 and 9). For this reason, the first analysis on the distance data was a multilevel gamma

regression (with a log link function). I ran this multilevel gamma regression with blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction predicting the Euclidean distance traveled every three frames (every tenth of a second). For all distance per frame analyses I aggregated the data by every three frames. This aggregation was necessary to reduce the significant runtime of subsequent Bayesian analyses due to their computational complexity (detailed below). The model results were back-transformed to the original data scale for Figure 8. All distance values of zero were transformed into half of the lowest non-zero distance value for the gamma regression. Gamma regressions cannot model 0 values, but it was important to retain all distance data provided by the participants. Random slope and intercept effects for the variables of video and participant were included in these analyses to allow for variability in the average distance measures and unique slopes across each video and participant (i.e., participants may show different sensitivity to the blur and window manipulations, and the movements recorded may differ uniquely depending on which video is presented). Multiple random effect structures were tested, and the best fitting model was found to be the one that included random slope effects of blur, window, and their interaction for videos and participants, and intercept effects for videos and participants. Model comparison is detailed more in Table 5.

Figure 7

Raw Distributions of MCBRD and DIEM Euclidean Distance Per Frame



Note. MCBRD (left, purple) and DIEM (right, teal) Euclidean distance measured in degrees of visual angle (dva). These values are before the aggregation by three that was used for analyses. These distributions are heavily right skewed (zoomed-in versions of these figures are shown in Figure 9) thus the initial choice to analyze the data with a multilevel gamma regression.

Table 5

Model Comparisons for Gamma Regressions Predicting Distance

Model	AIC	BIC
1. Intercept effect, participant: (1|participant)	-359,385	-359,312
2. Intercept effect, participant and video: (1|participant) + (1|video)	-390,748	-390,663
3. Blur and window slope and intercept effect, participant and video: (blur + window | participant) + (blur + window | video)	-397,942	-397,735
4. Blur, window, and interaction slope and intercept effect of participant and video: (blur * window | participant) + (blur * window | video)	-400,778	-400,475

Note. Models only differed in their random effects. The fixed effects of blur condition, window condition, and their interaction were kept the same for all models. The R syntax for each random effect structure is included. With the lowest AIC and BIC values Model 4 was chosen as the best fit.

The parameter estimates for the best fitting model are featured in Table 6. This model found a significant negative effect of window radius predicting distance traveled every three frames (B = -0.07, t = -4.34, p < .001) such that, as window radius increased, the predicted distance traveled decreased. The effect of blur approached significance (B = 0.16, t = 1.92, p = .054) suggesting that blur could have a positive effect with higher levels of blur resulting in more distance traveled on average. The estimated marginal means of this model are plotted in Figure 8 along with the DIEM baseline mean and standard error.

Table 6

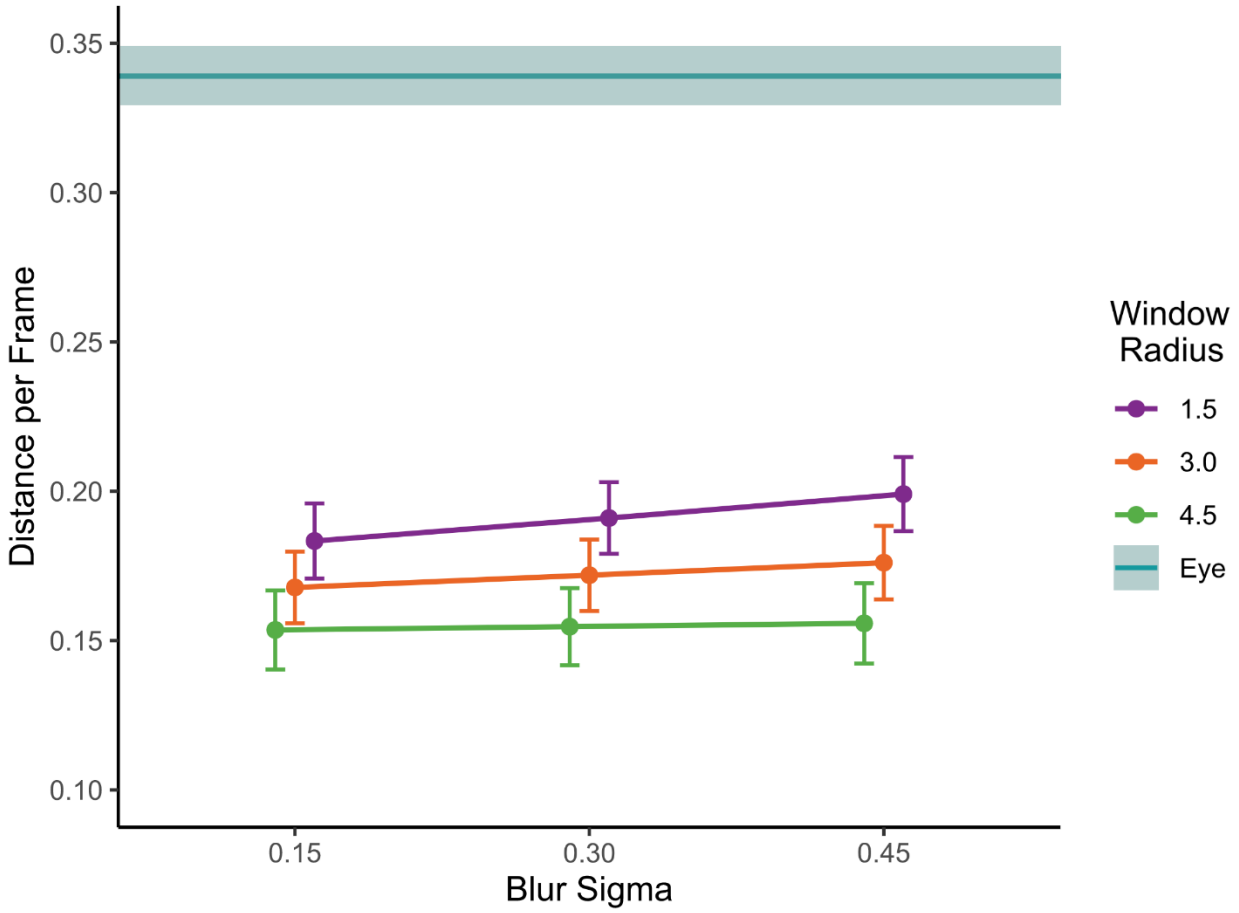
Parameter Estimates for Gamma Regression Predicting Distance

	B	*SE*	*t*	*p*
Intercept	-0.66	0.07	-9.51	< .001
Blur	0.16	0.08	1.92	.054
Window	-0.07	0.02	-4.34	<.001
Blur * Window	-0.07	0.07	-1.03	.303

Note. Multilevel gamma regression with blur sigma, window radius, and their interaction predicting the Euclidean distance traveled every three frames. This model included random intercept effects of video and participant, and random slope effects of blur, window, and their interaction for video and participant.

Figure 8

Estimated Marginal Means From Gamma Regression Predicting Distance



Note. Estimated marginal means from the multilevel gamma regression predicting distance (in dva) per three frames (back transformed from the original log scale of the model and back transformed to match the distance-per-frame scale the distance values were originally calculated on). Error bars are equal to 1 *SE*. The DIEM mean (0.34) is featured as a solid teal line with an error ribbon representing 1 *SE* (0.01). For information on how this baseline was generated see DIEM Baselines above.

Tukey's post-hoc contrast tests were run to investigate the differences between individual window and blur settings. These contrasts showed no significant differences between the outcome of blur levels at the small (1.5 dva), medium (3.0 dva), or large (4.5 dva) window radii. Contrast tests did find, however, that the outcomes of window radii did significantly differ from each other at each blur level. Overall, the smallest window size of 1.5 dva resulted in the most distance traveled. Despite these significant differences, all MCBRD conditions resulted in distance traveled per frame that was less than 75% that of the DIEM eye tracking data (less than 0.26 dva covered per frame). As detailed in the DIEM Baselines section above, this group distance means the MCBRD distances traveled per frame are shorter than the eye tracking distances on average in a way that undermines the importance of any value being significantly different and closer to the DIEM baseline (this determination is subjective, with the rationale that the MCBRD results are closer to half of the DIEM results). These values may still be used to influence MCBRD settings at a user's discretion, but the MCBRD paradigm is not recommended for any research that requires the rapid speed of eye movements (e.g., replicating saccade speed/distance effects or showing briefly-flashed images in a Rapid Serial Visual Presentation paradigm).

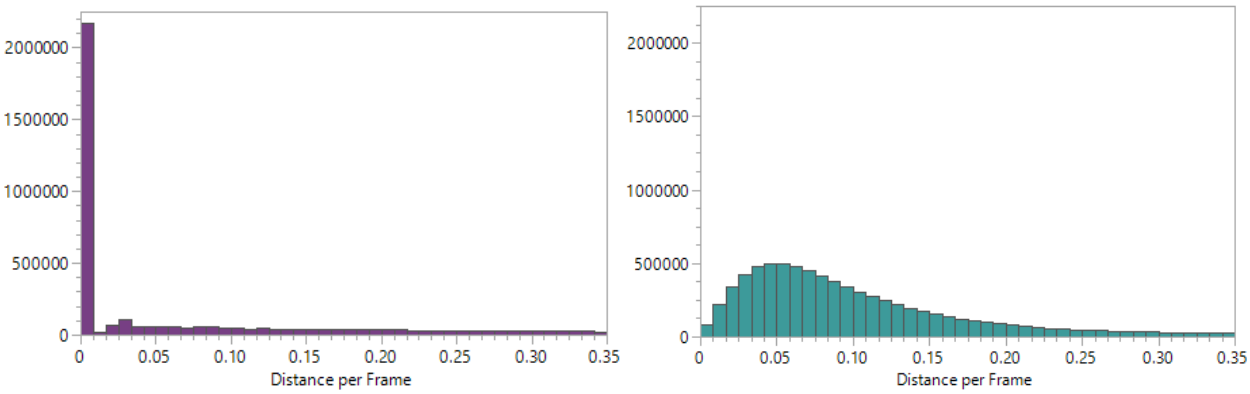
Bayesian Left-Censored Regression:

The distance-per-frame values from the MCBRD paradigm differed from the DIEM eye tracking data in one particularly interesting way. The MCBRD resulted in many more zero-distance frames than eye movements. This prevalence of zero-distance frames in mouse movements and not eye movements makes sense given the physical construction of the eye. Even when making a fixation, the eye moves slightly with microsaccades and drifting (Engbert & Kliegl, 2004). Mouse movements do not have this miniscule jitter and drift. Pauses in mouse

motion result in pure zero values for distance covered, and that results in different distributions of the raw distance data between mouse movements and eye movements (see Figure 9)

Figure 9

Zoomed-in Raw Distributions of MCBRD and DIEM Euclidean Distance per Frame



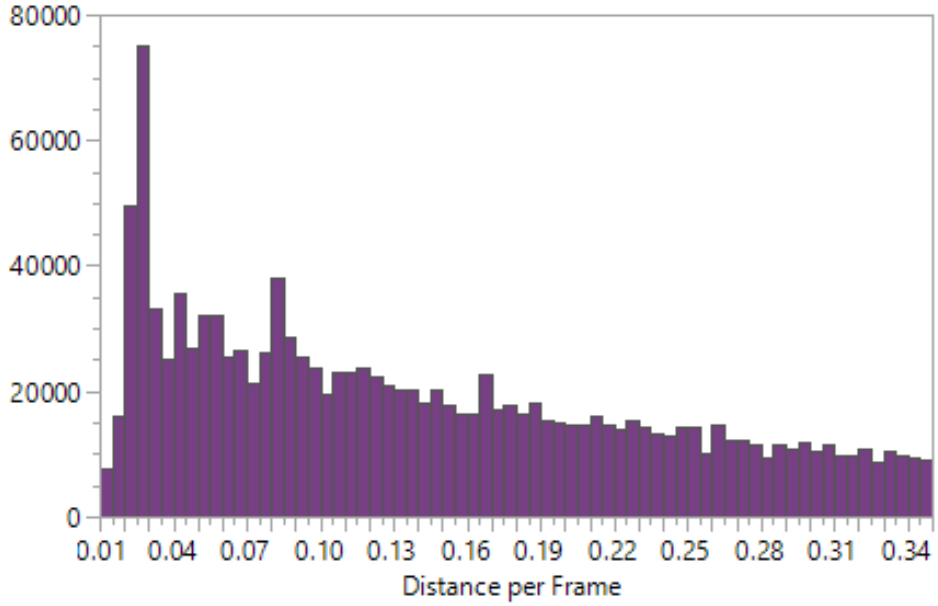
Note. Raw distributions of MCBRD (left, purple) and DIEM (right, teal) Euclidean distance per frame in degrees of visual angle (dva). These values are before the aggregation by three that was used for analyses. The MCBRD data had many zero values, and the DIEM data had a smoother distribution over small non-zero dva values.

The zoomed-in distribution of the MCBRD distance data points without zero values is shown in Figure 10. To account for this distribution, the distance analyses were also run as a left-censored regression. A left-censored regression allows a model to be fit that has unique variance at and below a censored point.

As the model would still be run as a gamma regression, the zero values were transformed to be equal to half of the smallest non-zero value. I treated this new value as the censor point. I used a gamma model because the distribution without the censored values (Figure 10) still showed a heavy positive skew.

Figure 10

Zoomed-in Raw Distributions of MCBRD Euclidean Distance Without Zeroes



Note. This figure shows the distribution of the non-zero MCBRD Euclidean distance per frame values in degrees of visual angle (dva).

Running a multilevel censored regression in R requires running the analysis as a Bayesian model. Thus, I ran a multilevel Bayesian left-censored gamma regression using the *brms* library (Bürkner, 2017) in R. As with the previous multilevel gamma regression, blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction were used as predictors of the Euclidean distance traveled every three frames (every tenth of a second). I aggregated the distance data to shrink the size of the dataset, to lessen the runtime of the tested Bayesian models. The large amount of data and complex random effect structure resulted in extremely long model-fitting times (running for two weeks in the longest case, even with multi-chain threading afforded by the cmdstanr backend, see

https://cran.r-project.org/web/packages/brms/vignettes/brms_threading.html for more information on *brm* threading). The aggregated data was used in the original gamma models for consistency with the Bayesian models.

Models with different random effect structures were tested similar to the previous gamma models. The results of this model comparison are detailed in Table 7. Here, as it was for the original gamma models, the best fitting model was found to be the one that included blur, window, and their interaction as terms that could vary in slope for individual videos and participants. I ran this model with 15,000 total iterations, with 1,500 burn-in iterations.. These iterations were run across 4 chains that were each threaded to run 16 computer cores each (running on 64 cores total). This Bayesian model included priors for the intercept value that followed a student's t distribution, with 3 degrees of freedom (default), -1 as the mean (a conservative prior that suggests an initial negative value for distance traveled per frame), and 4 as the variance. Both the random effects and regression weights of the model were given priors of a student's t distribution with 3 degrees of freedom, a mean of 0, and a variance of 3. Random effect correlations were given prior following a LKJ(2) distribution.

Table 7

WAIC Comparisons for Bayesian Left-Censored Regressions predicting Distance

Model	WAIC
1. Intercept effect, participant and video: (1|participant)+(1|video)	1,736,363
2. Blur and window slope and intercept effect, participant and video: (blur + window | participant) + (blur + window | video)	1,734,864
3. Blur, window, and interaction slope and intercept effect of participant and video: (blur * window | participant) + (blur * window | video)	1,734,296

Note. Models only differed in their random effects. The fixed effects of blur condition, window condition, and their interaction were kept the same for all models. The R syntax for each random effect structure is included. With the lowest WAIC value, Model 3 was chosen as the best fit.

The parameter estimates for this model are shown in Table 8, and the estimated marginal means of this model are in Figure 11. The results of this model are similar to those of the original gamma regression run on this data. In these results, only the effect of window radius does not have 0 within its 95% confidence intervals. This result suggests that it is unlikely that the effect of window radius on distance traveled per frame is zero. The similarities between the two models run for the distance-per-frame measure are further cemented when comparing Figures 8 and 11.

.

Table 8

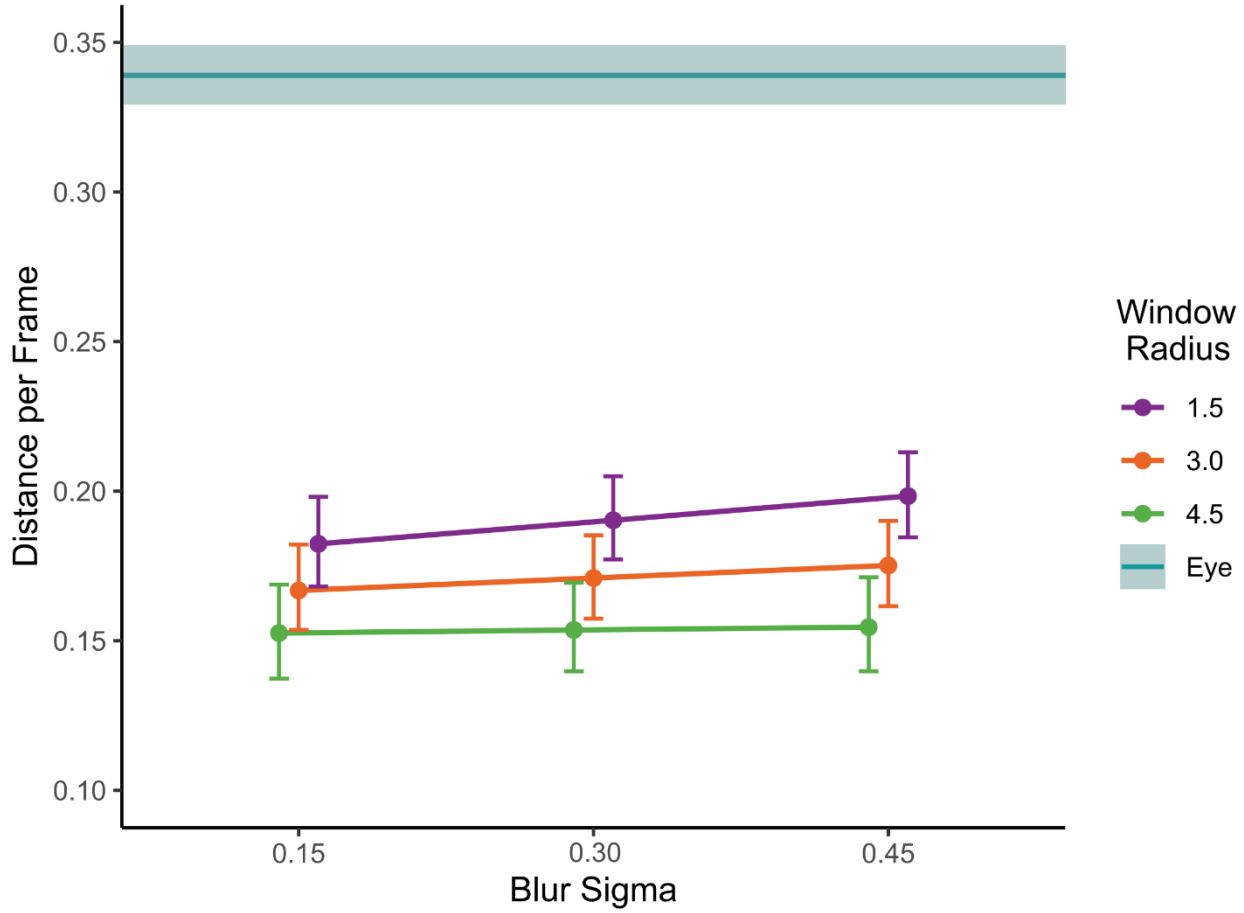
Parameter Estimates for Bayesian Left-Censored Regression Predicting Distance

	B	Est.Error	l-95% CI	U-95% CI	Rhat
Intercept	-0.67	0.08	-0.83	-0.51	1.001
Blur	0.16	0.13	-0.09	0.41	1.000
Window	-0.07	0.02	-0.11	-0.04	1.000
Blur * Window	-0.08	0.11	-0.29	0.14	1.000

Note. This model included random intercept effects of video and participant, and random slope effects of blur, window, and their interaction for video and participant.

Figure 11

Estimated Marginal Means From Bayesian Left-Censored Regression Predicting Distance



Note. Estimated marginal means from the multilevel Bayesian left-censored gamma regression predicting distance per three frames (back transformed from the original log scale of the model, and back transformed to match the distance-per-frame scale the distance values were originally calculated on). Error bars are equal to half of the distance from the mean estimate to the upper and lower HPD values reported by the Bayesian model. DIEM mean of 0.34 is featured as a solid teal line with an error ribbon representing 1 *SE* (0.01). For information on how this baseline was generated, see DIEM Baselines above.

Similar to the conclusions drawn from the earlier gamma regression these results are presented with the disclaimer that the MCBRD distance values are close to half of the distance values generated by eye movements. These results can be used to inform how distance and velocity may change with different window and blur levels, but these results suggest that no blur sigma and window radius will result in mouse movements that move at the speed and frequency of eye movements.

Normalized Scanpath Similarity

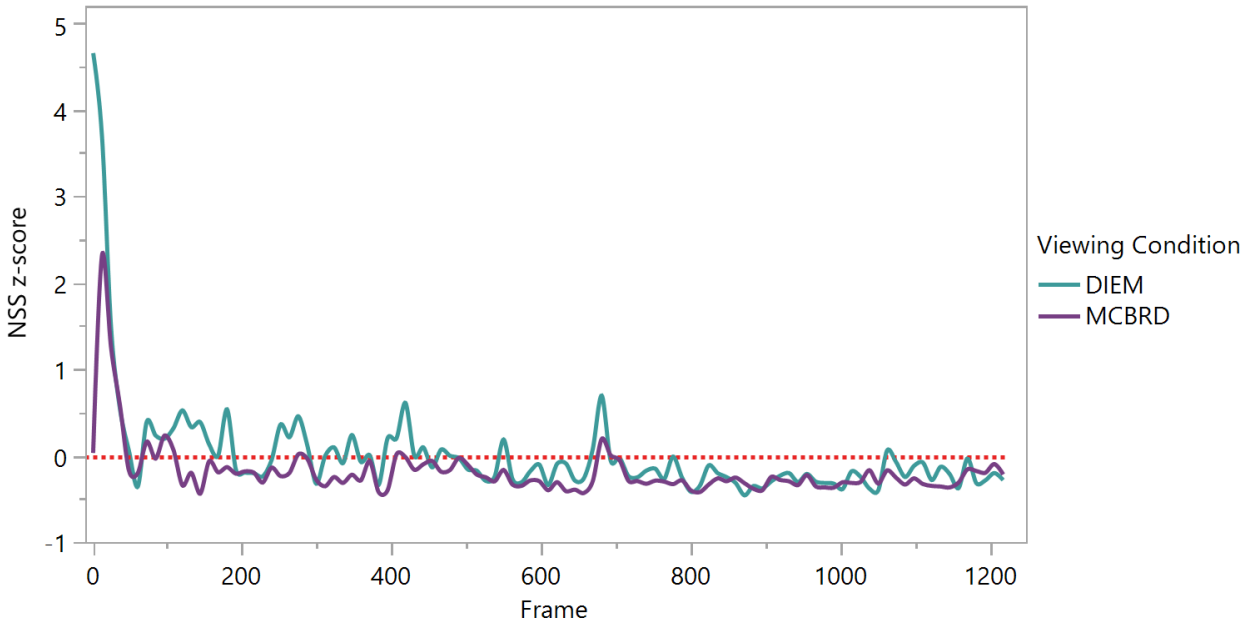
To analyze the similarity of attentional allocation between MCBRD and DIEM participants on the same videos, I used the Normalized Scanpath Saliency (NSS) method of comparing gaze heatmaps (maps of the probability of gaze position) calculated for each video frame (Bylinskii et al., 2018; Dorr et al., 2010; Le Meur & Baccino, 2013; Loschky et al., 2015; Peters et al., 2005). The implementation of this NSS method is similar to that used in Hutson et al., 2017. I used the DIEM eye tracking data, already downsampled to 30 Hz, to create the probability map of the reference condition. I placed a circular multivariate Gaussian distribution with a sigma (i.e., standard deviation) of 2 degrees (rough size of the fovea) around each gaze location and averaged over a temporal sigma of 280 ms (a moving window of time approximately the length of an average eye fixation). I summed and normalized the Gaussians using the mean and standard deviation of all DIEM participant values except one—this process was done in a leave-one-out procedure repeatedly for each DIEM participant so that each participant's gaze locations could be compared to the group without them in it. In these comparisons the z-score of similarity is equal to a participant's gaze location for one frame minus the mean from the summed Gaussians for that frame divided by the standard deviation of

that mean value. When the MCBRD values were compared to the DIEM probability map, the leave-one-out procedure was skipped.

The resulting output for each participant is a set of z-scores for each frame that represent their similarity to the Gaussian-smoothed probability map of the DIEM participants. Positive z-scores indicate higher gaze similarity than the mean, a z-score of zero represents an average similarity, and a negative z-score represents less similarity than the mean. These z-scores (comparing all MCBRD results to the DIEM probability map, and all DIEM results to the DIEM probability map using a leave-one-out procedure) were then used as the outcome variable for further analyses. For visualization of this measure the mean similarity results by frame for each group compared to the DIEM ground truth probability map for Video 1 are illustrated in Figure 12. The NSS z-score by frame figures for all other videos are featured in Appendix B.

Figure 12

Gaze Similarity of MCBRD and DIEM Participants for Video One



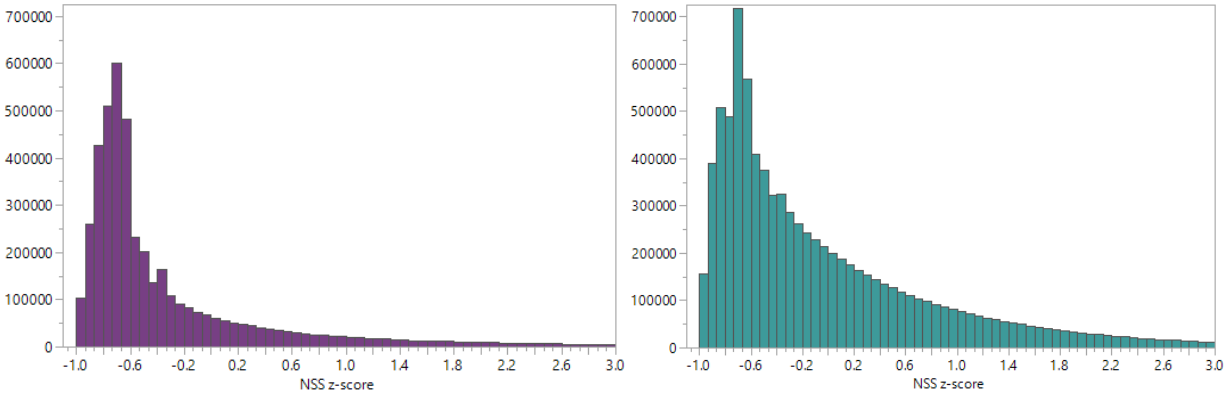
Note. Mean z-score of gaze similarity of MCBRD participants (in purple) and DIEM participants (in teal) compared to the DIEM reference gaze probability map for Video 1. DIEM participants were compared to a probability map made from all DIEM data but theirs in a leave-one-out method. A positive z-score represents greater similarity than average, a z-score of zero (indicated by a red dotted line) represents average similarity, and a negative z-score represents lower similarity than average. For NSS z-score by frame figures for all other videos see Appendix B. The videos for MCBRD participants were entirely blurred at the beginning, to prompt users to make the first mouse movement. Data recording for MCBRD participants began with the first mouse movement and was entered as NA values before that point. This resulted in lower NSS z-scores for MCBRD participants at the very beginning of each video.

The distributions of these z-scores are shown in Figure 13. Due to the heavy positive skew and the presence of negative values ending at -0.97, I transformed this data with a log(x +

1) transformation. The transformed distributions are shown in Figure 14. These transformed z-scores were used as the outcome variable in a multilevel linear model with blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction as predictors. The model results were back-transformed for the estimated marginal means plot in Figure 15. Multiple random effect structures were tested, and the best fitting model included random slope effects of blur, window, and their interaction for videos and participants, and intercept effects for videos and participants. Model comparison is featured in Table 9.

Figure 13

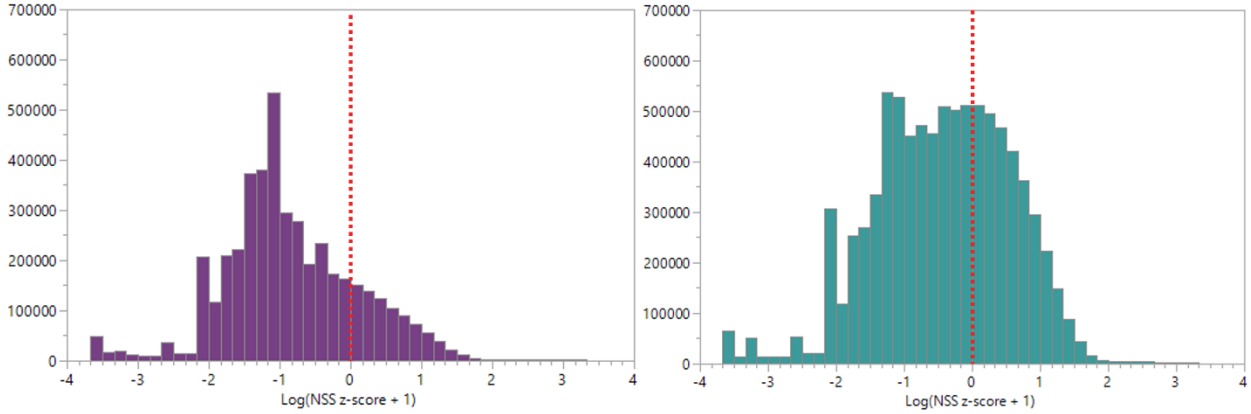
Raw Distributions of MCBRD and DIEM NSS Z-Scores



Note. Raw distributions of MCBRD (left, purple) and DIEM (right, teal) NSS z-scores. These distributions are heavily right skewed and were transformed by log(x + 1) before being used in a multilevel linear regression.

Figure 14

Distributions of MCBRD and DIEM NSS Z-Scores After Log(x + 1) Transformation



Note. Distributions of MCBRD (left, purple) and DIEM (right, teal) NSS z-scores after log(x + 1) transformation. A vertical red line is included at x = 0 for comparison between the two distributions.

Table 9

Model Comparisons for Linear Regressions Predicting Gaze Similarity

Model	AIC	BIC
1. Intercept effect, participant: (1|participant)	11,335,582	11,335,662
2. Intercept effect, participant and video: (1|participant)+(1|video)	10,981,181	10,981,274
3. Blur and window slope and intercept effect, participant and video: (blur + window | participant) + (blur + window | video)	10,919,190	10,919,416
4. Blur, window, and interaction slope and intercept effect for participant and video: (blur * window | participant) + (blur * window | video)	10,908,519	10,908,851

Note. Model comparisons for multilevel linear regressions predicting log(NSS z-score + 1). Models only differed in their random effects. The fixed effects of blur condition, window condition, and their interaction were kept the same for all models. The R syntax for each random effect structure is included. With the lowest AIC and BIC values Model 4 was chosen as the best fit.

The best fitting model did encounter convergence errors but the model parameters were found to be similar across different *lme4* optimizers (including the default nloptwrap optimizer and the other six optimizers included in the *allfit* function in the lme4 library; see the allfit() section of the lme4 manual at https://cran.r-project.org/web/packages/lme4/lme4.pdf for more information). The lme4 manual suggests that if model parameters are practically equivalent between tests with different optimizers, convergence warnings can be considered to be false positives. Furthermore, the final optimizer chosen for this model, NLOPT_LN_PRAXIS, resulted in the model coming closer to convergence with max|grad| = 0.0028 (the tolerance of this model convergence was set to max|grad| = 0.002). Finally, the model fixed effect beta weights were similar to those of the other models with less complex random effects structures

(models 1, 2, and 3 in Table 9). For these reasons, I am comfortable reporting Model 4 as the best fitting model and describing the results below.

The parameter estimates for the best fitting model are featured in Table 10. The model found significant effects of blur sigma (B = 0.28, t = 3.18, p < .01) and window radius (B = -0.03, t = -2.43, p = .02) predicting participants similarity to the DIEM probability map in terms of their NSS z-scores (with z-scores log(x+1) transformed). The effect of the blur sigma was positive such that as blur sigma values increased, the predicted NSS z-score increased. The effect of window size was negative such that as window size increased, the NSS z-score decreased. The estimated marginal means of this model are shown in Figure 15, zoomed out to show the DIEM baseline, and in Figure 16, zoomed in to show differences between the 3 × 3 MCBRD conditions.

Table 10

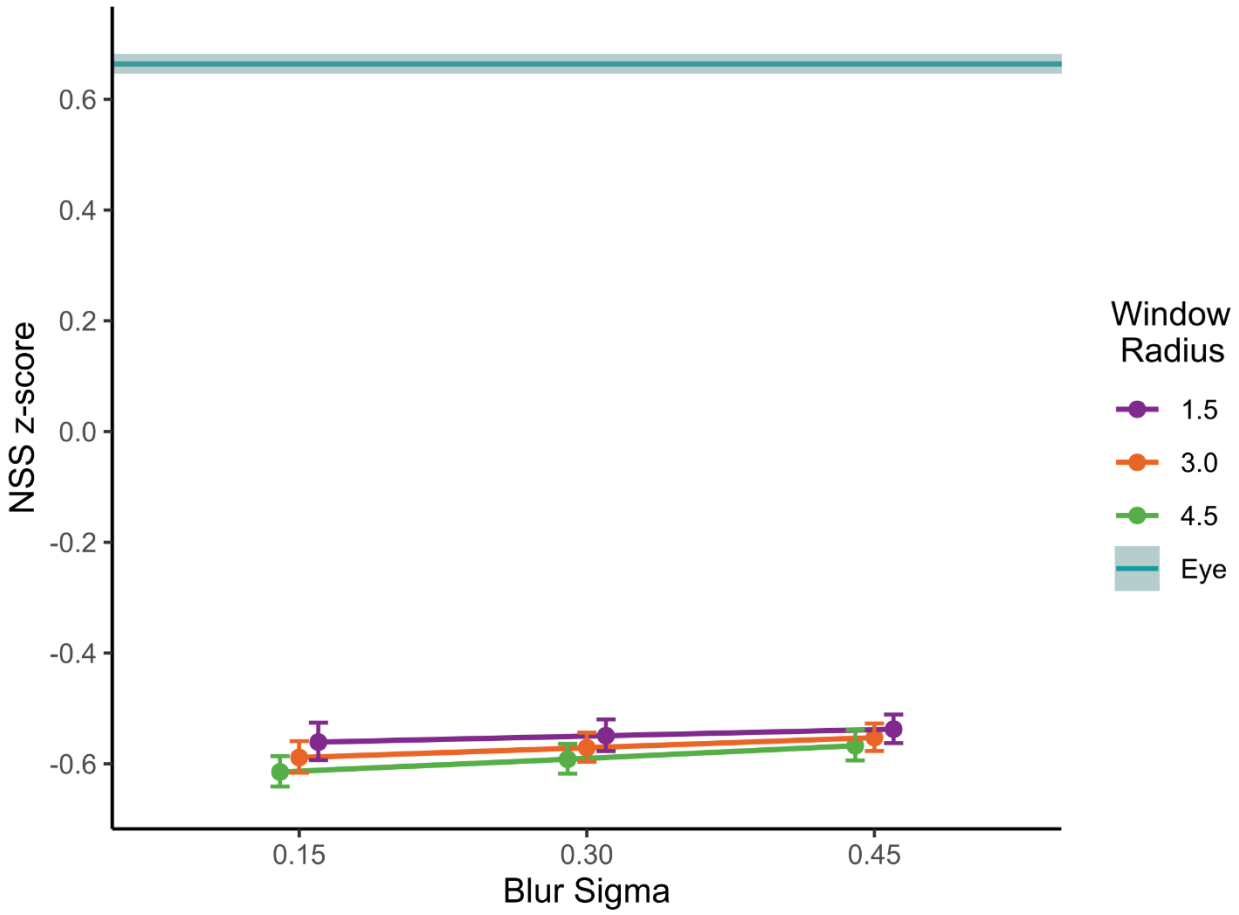
Parameter Estimates for Linear Regression Predicting Gaze Similarity

	B	*SE*	*df*	*t*	*p*
Intercept	-0.85	0.06	36.72	-13.82	<.001
Blur	0.28	0.09	68.52	3.18	.002
Window	-0.03	0.01	39.85	-2.43	.020
Blur * Window	0.07	0.06	46.38	1.18	.245

Note. Parameter estimates for the multilevel linear regression predicting log(NSS z-score + 1). This model included random intercept effects of video and participant, and random slope effects of blur, window, and their interaction for video and participant. Estimates are given on the scale of log(x+1).

Figure 15

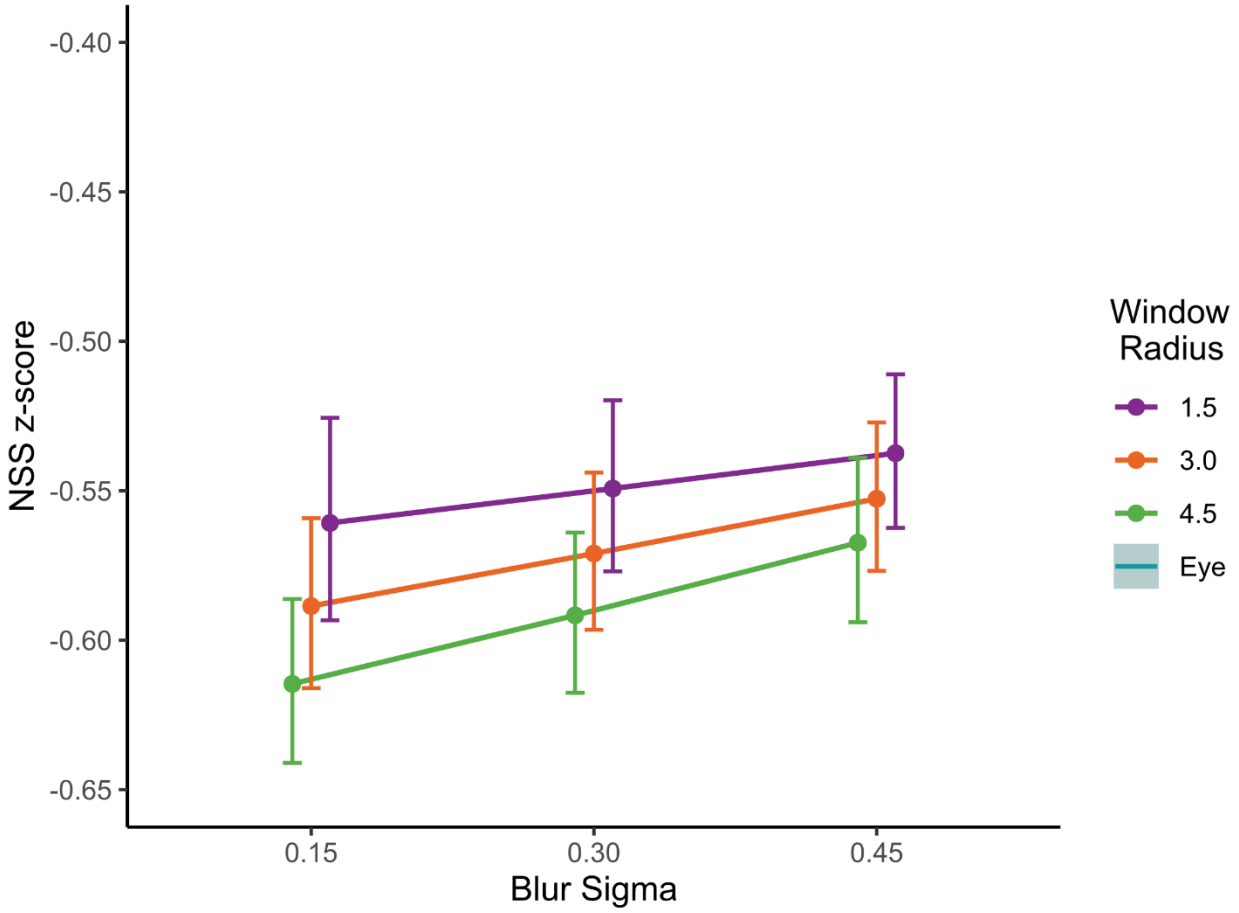
Estimated Marginal Means From Linear Regression Predicting Gaze Similarity



Note. Estimated marginal means from the multilevel linear regression predicting log(NSS z-score + 1) . Values are back-transformed for this figure. Error bars are equal to 1 *SE*. DIEM mean (0.66) is featured as a solid teal line with an error ribbon representing 1 *SE* (0.02). For information on how this baseline was generated see DIEM Baselines above.

Figure 16

Zoomed-In Estimated Marginal Means From Linear Regression Predicting Gaze Similarity



Note. Estimated marginal means from the multilevel linear regression predicting log(NSS z-score + 1). Values are back-transformed for this figure. Error bars are equal to 1 *SE*.

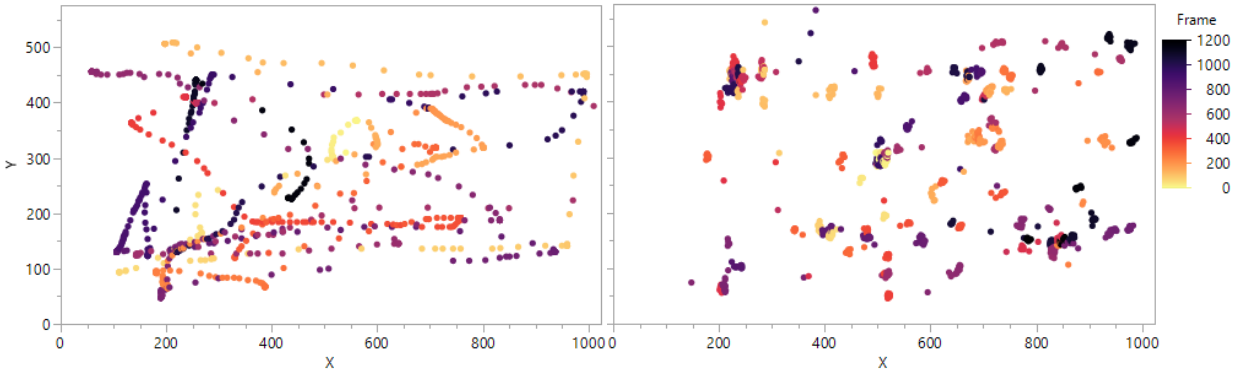
Tukey's post-hoc contrast tests were run to determine differences between NSS z-scores under different blur sigma and window radius conditions. These contrasts showed a significant effect of blur when the window radius was 3 dva (*B* between conditions = -0.0418, *z* = -3.180, *p* < .01) or 4.5 dva (*B* = -0.0578, *z* = -3.739, *p* < .001). Contrasts also found a significant effect of window radii, but only at blur sigma values of 0.15 dva (*B* = 0.065, *z* = 2.44, *p* = .04) or 0.30 dva (*B* = 0.049, *z* = 2.43, *p* < .001). However, an important finding here is that all MCBRD setting

conditions resulted in negative NSS z-scores. These negative z-scores mean that the MCBRD x-y coordinates on each frame were more different than the DIEM ground truth probability map than average, when compared to how the DIEM eye tracking participants' data correlated to the ground truth map. As described in the DIEM Baselines section above, this difference diminishes the importance of significantly different values in the MCBRD 3 × 3 conditions. On average, MCBRD movements differ from eye movements. I believe this movement difference to be due to the inherent differences between hand movements and eye movements, discussed earlier when describing the similar influence of this factor on the distance per frame measures.

This difference is easier to grasp when visualizing the scanpaths between a mouse tracking participant and an eye tracking participant. Example scanpaths for both are plotted in Figure 17. Although this figure only shows data for one participant from each group, it is characteristic of the data found for all participants in the two viewing paradigms. Within Figure 17, despite the two participants attending to similar regions of the screen at similar times, their motion patterns differ. Specifically, the mouse movements are slow enough that points along the motion paths are captured at the downsampled 30 Hz data rate. The eye movement data, also downsampled to 30 Hz, reflects movements fast enough that the motion paths are difficult to distinguish in the plotted data points. NSS analyses, comparing data points on a frame-by-frame basis, would be sensitive to these motion differences.

Figure 17

Raw Data Points for MCBRD Participant One and DIEM Participant One Viewing Video One



Note. Raw data points for MCBRD Participant 1 (left) and DIEM Participant 1 (right) viewing Video 1. The data points are colored with a gradient ranging from yellow to dark purple, with yellow points representing the early frames of the video, and dark purple representing the later video frames. The slower motion of the mouse movements (left) is perceptible as distinct motion paths in the 30 Hz data.

The final analyses will examine whether, despite the differences in types of motions made between the hand and the eye, participants viewing videos in both paradigms attended to semantically important Areas of Interest at a similar rate.

Areas of Interest Analyses

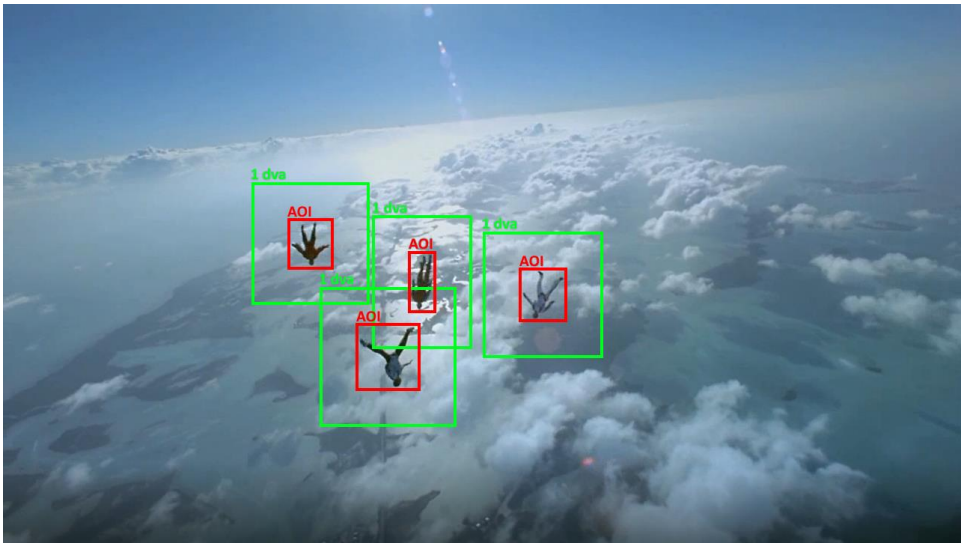
Dynamic Areas of Interest were created to capture semantically important characters and objects in 18 videos. These characters and objects were manually identified in rectangular regions and tracked along subsequent frames using the Dynamic Regions Tracker (DRT) program (Bonikowski et al., 2021) using the Kernelized Correlation Filters (KCF) object tracking method. AOIs were not created for videos without semantically important characters or

objects (like Video 26, time-lapse footage of a building being constructed) or videos with too many rapid cuts that introduced difficulties with the object tracking software.

The DRT program produces an output with the x-y coordinates of AOIs across all frames, including the top y value and leftmost x value, alongside a height and width measure in pixels. I wrote custom code in R to parse through the MCBRD and DIEM x-y coordinate data for each frame to determine whether a participant's coordinates were within an AOI with 1 dva of error (a standard amount of error found in eye trackers that is commonly used as an error value in AOI analyses, see Holmqvist & Andersson, 2017) on any given frame. AOIs were created with rectangle borders close to the edges of important characters or objects to ensure that the 1 dva margin of error did not result in too large of AOI regions. A visualization of AOIs created and the 1 dva error regions is shown in Figure 18.

Figure 18

Example AOIs With One Degree of Visual Angle Error Regions Around Each



Note. Example AOIs are shown in red, and error regions of 1 dva are shown in green.

I analyzed the total time spent within AOIs throughout video frames that had AOIs present. Total time is used as the predicted output in this analysis instead of the common AOI analysis of total fixation time because there is currently no distinction made between saccades and fixations in mouse movement data (see the Future Analyses section for suggestions on how this distinction could be made). With the slower rate that participants moved their mouse-contingent window compared to eye movements (see Distance per Frame analyses above), visual information can also be collected from a moving mouse-contingent window, as opposed to eye saccades that prevent the input of visual information through saccadic suppression. For this reason, I consider the total time spent in AOIs to be the more appropriate analysis.

Instead of the common method of comparing AOI results with *t* tests of the proportion of fixations on selected AOIs between conditions (Hutson et al., 2017; Vo et al., 2012 as examples), I analyzed this data with a multilevel binomial logistic regression. I formatted the data for each participant with each possible AOI frame receiving either a 0 if the participant's coordinates were not within an AOI plus error region in this frame (considered a *miss*), or a 1 if they were (considered a *hit*). I ran a multilevel binomial logistic regression with blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction predicting the probability of hits on each frame. I included random slope effects of blur, window, and their interaction for videos and participants, and intercept effects for videos and participants in this analysis. I chose this random effect structure from multiple tested models detailed in Table 11. Similar to the reported NSS model above, I ran this model with the NLOPT_LN_PRAXIS optimizer and the model converged.

Table 11

Model Comparisons for Binomial Logistic Regressions Predicting AOI Hit Probability

Model	AIC	BIC
1. Intercept effect, participant: (1|participant)	1,699,826	1,699,888
2. Intercept effect, participant and video: (1|participant)+(1|video)	1,606,352	1,606,427
3. Blur and window slope and intercept effect, participant and video: (blur + window | participant) + (blur + window | video)	1,594,105	1,594,304
4. Blur, window, and interaction slope and intercept effect for participant and video: (blur * window | participant) + (blur * window | video)	1,589,159	1,589,457

Note. Model comparisons for multilevel binomial logistic regressions predicting AOI hits on each frame. Models only differed in their random effects. The fixed effects of blur condition, window condition, and their interaction were kept the same for all models. The R syntax for each random effect structure is included. With the lowest AIC and BIC values Model 4 was chosen as the best fit.

The parameter estimates for the best fitting model are featured in Table 12. This model found a significant main effect of blur predicting the likelihood of AOI hits, with higher blur sigma values increasing the likelihood of AOI hits ($B = 0.60$, $z = 10.66$, $p < .001$). This model also found a significant interaction between blur sigmas and window radii in predicting AOI hits, with smaller window sizes predicting more AOI hits for the lowest blur sigma of 0.15 dva, and larger window sizes predicting increasingly higher rates of AOI hits for the medium blur sigma (0.3 dva) and highest blur sigma (0.45 dva) conditions ($B = 0.20$, $z = 5.52$, $p < .001$). This interaction can be visualized with the estimated marginal means plot in Figure 19 alongside the DIEM baseline.

Table 12

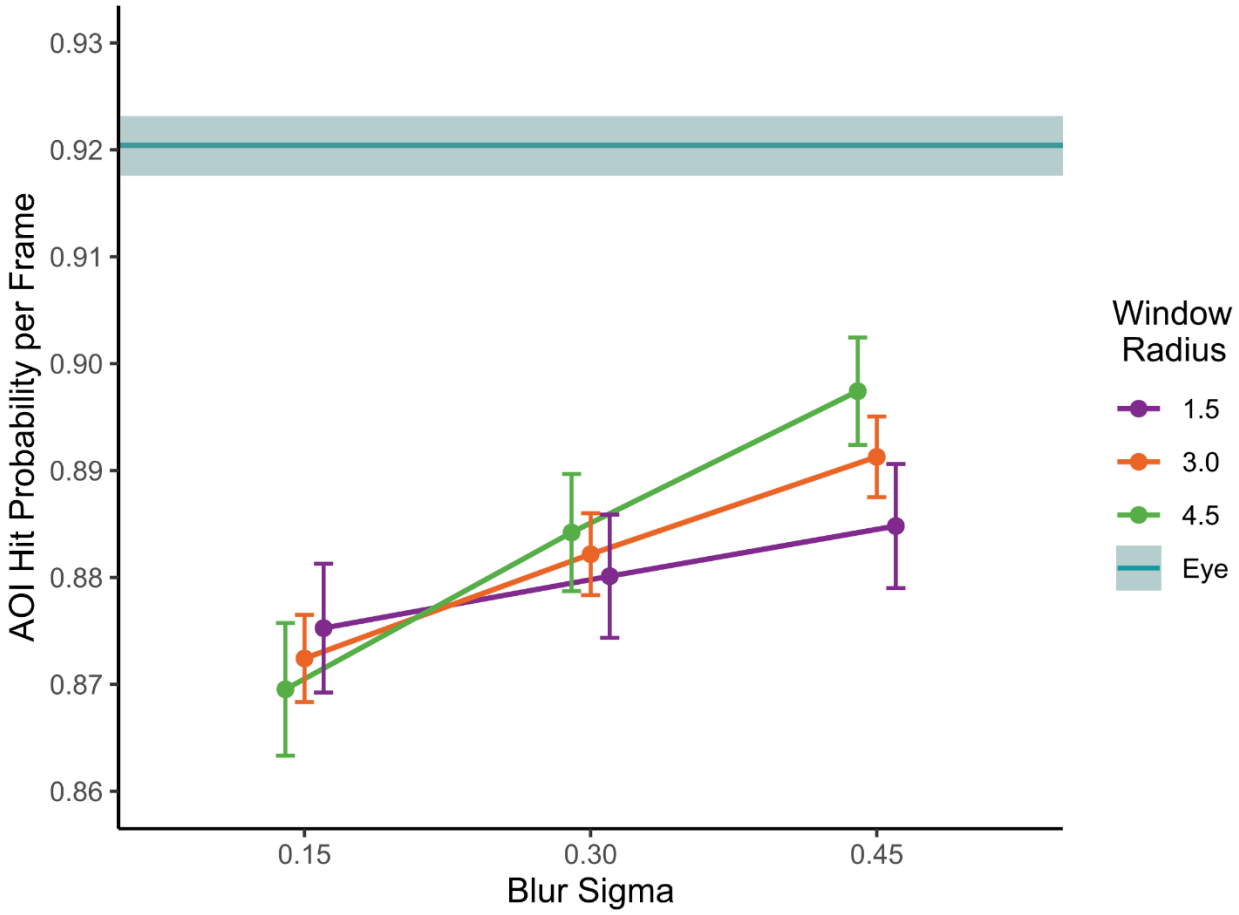
Parameter Estimates for Binomial Logistic Regression Predicting AOI Hit Probability

	B	*SE*	*z*	*p*
Intercept	2.01	0.04	54.64	<.001
Blur	0.60	0.06	10.66	<.001
Window	0.01	0.03	0.50	.619
Blur * Window	0.20	0.04	5.52	<.001

Note. Parameter estimates for the multilevel binomial logistic regression predicting AOI hit probability on individual video frames. This model included random intercept effects of video and participant, and random slope effects of blur, window, and their interaction for video and participant

Figure 19

Estimated Marginal Means From Binomial Logistic Regression Predicting AOI Hit Probability



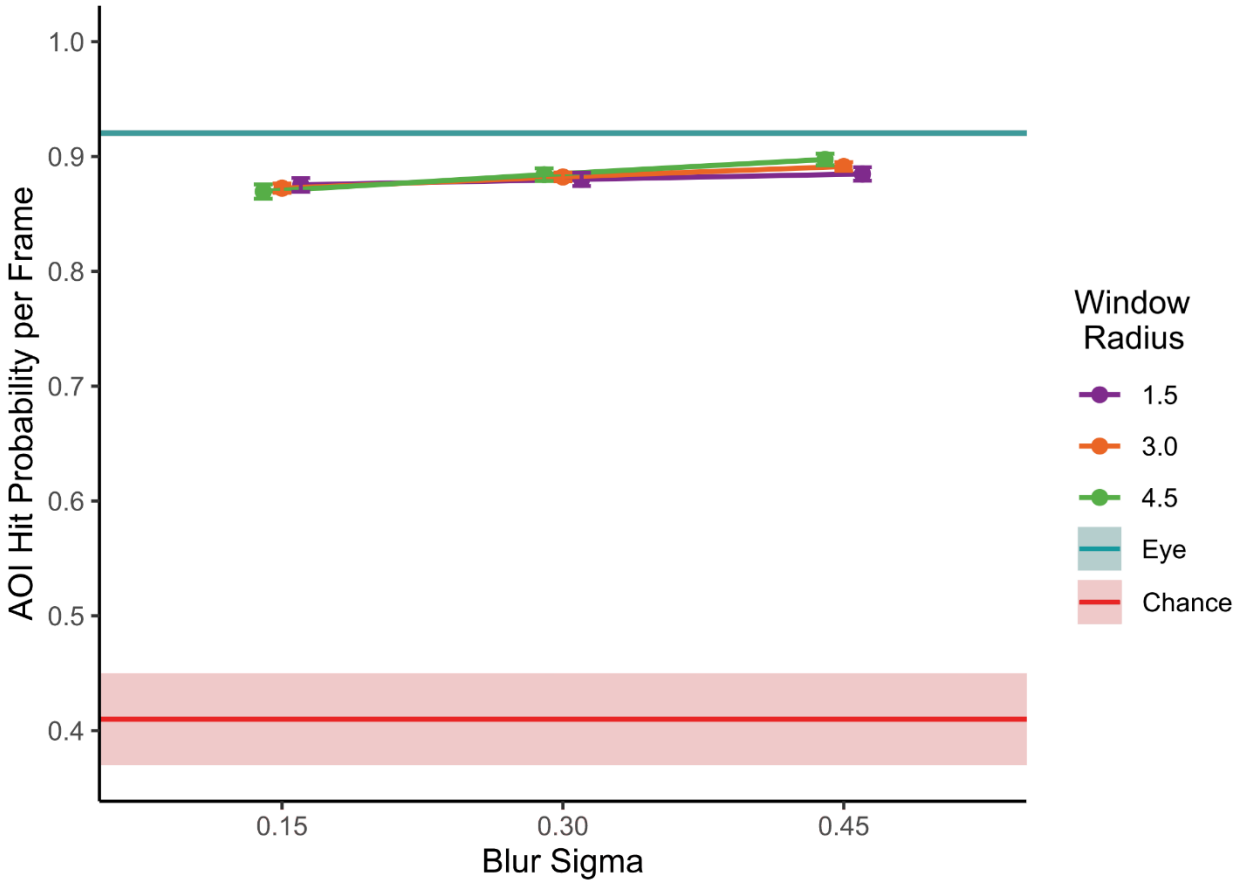
Note. Estimated marginal means from the multilevel binomial logistic regression predicting AOI hit probability on each frame. Error bars are equal to 1 *SE*. All values are back transformed from the logit scale of the original model. DIEM mean (0.92) is featured as a solid teal line with an error ribbon representing 1 *SE* (0.003). For information on how this baseline was generated, see DIEM Baselines above.

The MCBRD paradigm is comparatively much closer to eye tracking in terms of AOI results relative to the previous Distance per Frame and NSS analyses. The high similarity of the

MCBRD results to the DIEM baseline is noteworthy when compared to chance. I calculated chance as the percentage of pixels in AOI regions (with 1 dva margins of error on all sides) divided by the total number of pixels in each frame. The mean chance, averaged over the chance values calculated for all videos, was 0.41 (*SD* = 0.04). For visualization, this chance value is shown as the lower y axis limit in Figure 20.

Figure 20

Zoomed-Out EMMs From Binomial Logistic Regression Predicting AOI Hit Probability

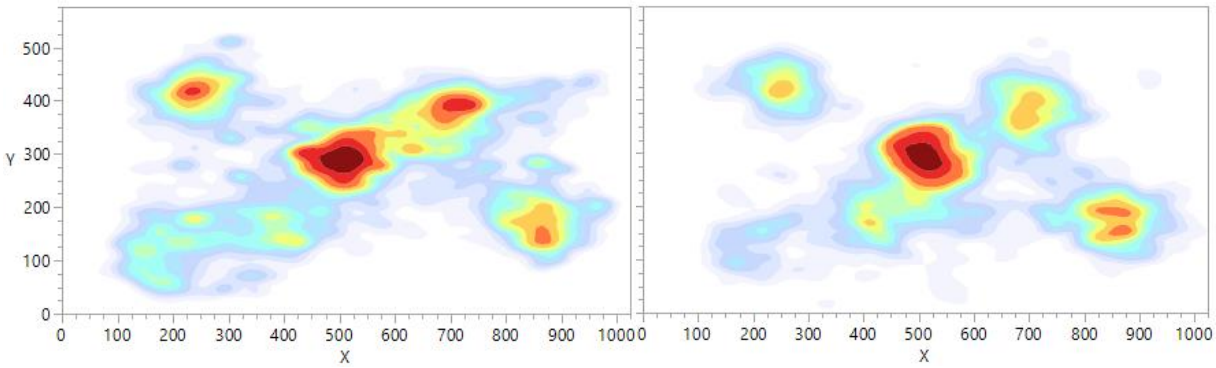


Note. Estimated marginal means from the multilevel binomial logistic regression predicting AOI hit probability on each frame. The Y axis minimum is the mean probability of AOI hits (including 1 degree of error) for any pixel on any frame, averaged over all AOI videos, $M = 0.41$, $SE = 0.04$. Error bars are equal to 1 *SE*. All values are back transformed from the logit scale of the original model. DIEM mean (0.92) is featured as a solid teal line with an error ribbon representing 1 *SE* (0.003). For information on how this baseline was generated, see DIEM Baselines above.

These results suggest that MCBRD and DIEM participants are both frequently attending to regions of semantic interest on the screen despite the motion differences between hand and eye movements. This conclusion is supported by an example heatmap of total x-y samples taken from one video in Figure 21, and the heatmaps for the rest of the videos shown in Appendix A.

Figure 21

Heatmaps From MCBRD and DIEM X-Y Coordinate Data From Video One



A. B.

Note. A: Heatmap generated from MCBRD x-y coordinate data (normalized to the scale of the original DIEM video presentation) from Video 1. B: Heatmap generated from DIEM x-y coordinate data from video 1. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions. As shown in the AOI analyses, participants viewing videos in MCBRD and eye tracking paradigms attended to similar regions of interest.

Due to this similarity in AOI hits to the DIEM data, the results found for the 3 × 3 MCBRD conditions are worth special consideration. Because the goal of these comparisons was

to identify the conditions that produce results closest to the DIEM baseline, these AOI comparisons are more valuable when the DIEM baseline is within reach.

Tukey's post-hoc contrast tests showed that the effect of window size was not significant at any blur sigma value (at the blur sigma of .15 dva, *B* between conditions (as an odds ratio) = 1.03, $z = 0.63$, $p = 0.80$; at the blur sigma of .3 dva, $B = 0.98$, $z = -0.50$, $p = 0.87$; at the blur sigma of .45 dva, $B = 0.94$, $z = -1.63$, $p = 0.23$). However, the effect of blur sigma was significant at any window size. At the window radius of 1.5 dva, *B* between conditions = 0.96, $z = -3.66$, $p < .001$. At the 3.0 dva window radius, $B = 0.91$, $z = -10.66$, $p < .001$. At the 4.5 dva window radius, $B = 0.87$, $z = -12.06$, $p < .001$. The results are qualified by the significant interaction described above. Furthermore, the *SE* of all MCBRD estimated marginal means (mean *SE* = 0.005, *SD* of this mean = 0.001) is larger than the *SE* of the DIEM baseline (0.003). Following these larger *SE* values, any significant results from this multilevel binomial logistic regression are deemed significant with a larger 95% confidence interval (2 * *SE* from the mean in the positive and negative direction) than that of the eye tracking data, so the differences are meaningful in that they are greater than the natural differences that would be found in eye tracking data.

These contrast tests found the highest blur sigma of 0.45 dva to produce AOI hit rates closest to those of eye movements. For this reason, a blur sigma of 0.45 is the recommended setting for MCBRD paradigms used for video stimuli. The difference between window radii was not significant at this blur sigma. One could choose to go with a window radius of 4.5 dva, as it resulted in the highest estimated marginal mean shown in Figure 19, or a lower window radius like 1.5 dva that elicited higher scores for distance traveled per frame (see Distance per Frame measures above). The 3.0 dva window radius is a good compromise between the two. However,

these recommendations are given with the caveat that the differences between AOI hit percentages are only as large as 2.7% in the most extreme case. Though the different blur settings result in significant differences, some users may find that a 2.7% increase does not warrant a hard rule on blur sigma level.

Chapter 4 - Discussion

This validation study did not definitively find optimal choices for blur sigma and window radius settings to use in a mouse-contingent bi-resolution display of video stimuli, but it does clarify how this paradigm could be used in the measurement of attention to video stimuli. I ran this validation with the goal of obtaining eye-like movements. For some of the observed results, reaching eye-like movements is simply unfeasible. In particular, Euclidean distance traveled per frame and gaze similarity to eye movements both suffered as a result of the motion differences between eye movements and hand movements. Eye tracking results for the 27 videos used here showed that participants moved their eyes an average of 0.34 degrees of visual angle (dva) per 30 Hz frame (*SE* = 0.01), whereas participants viewing videos through the MCBRD moved their high-resolution window about half this distance per frame, with an average of 0.17 dva (*SE* = 0.01). Eye tracking participants moved their eyes faster and more often than MCBRD participants moved their mice, and this result matches my expectation about the relative speed of both motion types.

My first set of competing hypotheses addressed whether or not eye tracking and MCBRD participants would generally attend to the same visual content even if their motion patterns differed. The results of this validation study suggest that participants do attend to similar content. With Areas of Interest created for different people and objects of semantic importance, I found that MCBRD AOI hit percentages (ranging from 87%, *SE* = 0.6% to 90%, *SE* = 0.5%) were closer on average to the eye movement reference baseline (*M* = 92%, *SE* = 0.3%) than they were to chance (*M* = 41%, *SD* = 4%). Furthermore, increases in blur sigma up to 0.45 dva were found to increase the AOI hit percentage ($B = 0.6$, $z = 10.66$, $p < .001$). For this reason, my primary recommendation is to use a blur sigma of 0.45 dva when setting up a MCBRD paradigm with

video. Because window radius was not found to have a significant effect at this blur level, I am tentatively recommending the middle value radius of 3.0 dva (for more on this recommendation, see the Areas of Interest Analysis section of the Results).

The second set of competing hypotheses concerned whether or not an optimal blur sigma and window radius setting would apply to all videos. Each of the final best-fitting models for the analyses run in this study included a random slope and intercept effect of video. These random effects accounted for individual variance in videos and how that variance may impact the effects of blur sigma, window radius, and their interaction. The effect of blur found in the AOI analysis occurs even when the variance from videos is taken into account. Thus, on average, the higher blur level of 0.45 dva can be expected to increase the AOI hit rate in a MCBRD paradigm.

This study did find significant effects of smaller window sizes predicting more distance traveled per frame, and of smaller window sizes and higher blur sigmas predicting slightly higher gaze similarity to eye tracking data. However, due to the larger distance between MCBRD values and the DIEM baseline in these analyses, it may not be fruitful to base MCBRD settings off of these significant differences. Even with the best settings found here, the distance and speed of mouse movements will not reach that of eye movements, and motion patterns will differ on average. I instead suggest a different role that a MCBRD paradigm could play in the measurement of attention to video.

This validation study shows that a Mouse Contingent Bi-Resolution Display can be used as a measure of attention to Areas of Interest in videos. I believe this separate measure of attention is the optimal role of this paradigm going forward. MCBRD paradigms are not a replacement for eye tracking, but they can be useful for determining semantically important visual content. These methods can be especially helpful in conditions where eye tracking is not

available (due to cost for example), or if the logistics of eye tracking are not optimal (if a researcher wants to collect a large sample without running participants one-at-a-time with an eye tracker). A researcher may find a mouse-contingent display suits their needs better than webcam-based eye tracking if they are concerned about eye tracking calibration errors, imprecision in measurements, or attrition due to technical issues or privacy concerns. Furthermore, this mouse contingent display paradigm is easy to implement into existing online experiment creation software and is easier to set up than eye tracking.

Limitations

One practical limitation of this study is its strict experimental control, which may reduce ease of use. Screen size and distance from the screen were measured to create blur levels and window radii that were based on degrees of visual angle in the participant's visual field. However, this visual angle calculation was a necessary step in this validation study and will be a necessary step in other studies where lab-quality data is desired. Despite this, some users may find these steps take away from the ease and speed of tracking mouse movements. To see if there is a way to extend the findings of this controlled validation even to people who want a simple mouse-contingent paradigm without the dynamic screen scaling measure, I intend to analyze the proportion of pixels used in the 0.45 dva blur condition and the 0.30 dva window condition out of the total screen size for all participants. If there is a general proportion of screen size that can be used instead to approximate these dva blur settings when presented at an average viewing distance, it may be valuable to add that as an option for those who would otherwise pick an uninformed screen proportion value instead of implementing degree of visual angle measurements.

In contrast to this, there were areas of the study where more control could have been beneficial. Particularly, the distance from the viewer to the screen was only measured once throughout the experiment. Viewers were instructed to stay in the same viewing position for the duration of the study, but additional distance measurements could have ensured that participants followed this instruction without moving in their seat or slouching to a degree that may have impacted the size of the video presentation in their field of view. A follow-up in-lab study could determine whether the findings from this project replicate when viewing distance is strictly controlled and participants are seated in a head rest.

Another limitation of this study is that there were specific analyses that could not validly be run but would have been of interest to readers. Within this study, the DIEM eye tracking data was only used as a baseline when interpreting the MCBRD 3 × 3 results. Because the DIEM eye tracking data was collected from an entirely different sample of participants, at a different location and at a different time, these two viewing conditions could not be reasonably compared in a statistical analysis. A future study could address this limitation by randomly assigning participants to an eye tracking or mouse tracking condition.

This study could not analyze the effects of computer setup on mouse movement patterns due to the overall sample size and the uneven number of participants responding with different viewing conditions (see Table 3). These variables are still important to consider and are worth addressing in a follow-up study (see the Future Analyses and Directions section below).

Finally, this study did not include a formal survey addressing how participants felt about each of the blur levels and window sizes. The ending survey did ask if participants were confused by anything in the study and if they had general feedback, but only three participants had feedback that mentioned any particular conditions by name. These three participants

reported not liking the smallest window (1.5 dva). A follow-up should be conducted with a more detailed user experience survey to determine if any of the tested conditions were more pleasant to work with.

Future Analyses and Directions

Future work can help to clarify the best ways to use this paradigm and to analyze the data from it. Some researchers may prefer a way to analyze mouse movement results similar to how they analyze eye movement results. Eye movements are generally analyzed by making the distinctions between saccades, fixations, and smooth pursuit movements. These distinctions are generally made in two ways: dispersion-based methods can identify fixations and non-fixations by identifying when the distance traveled surpasses a threshold. Velocity-based methods use velocity cutoffs to identify saccades and fixations and will sometimes implement acceleration thresholds to separate out smooth pursuit movements (Holmqvist & Andersson, 2017). These analyses were considered for this project, but I decided that these analyses would be best suited for a follow-up study. The threshold values that best separate fixation, saccade, and smooth pursuit movements for mouse movement data are unknown. Trial fixation analyses were conducted for data visualization purposes, and these analyses showcase the differences in movement types between eye and mouse movement data when measured with the same threshold. These differences, shown in Appendix D, suggest that it may be worthwhile to determine new thresholds to describe the different motions made by mouse movements. For these trial analyses I wrote a fixation detection algorithm in R to identify fixations as any time period where the participant's measured x-y coordinates move no more than 1 dva for a minimum duration of 100 ms (common thresholds for fixation detection, see Holmqvist & Andersson, 2017). Thresholds best suited for mouse movement data could be identified through

a cluster analysis, like Latent Class Analysis, using distance traveled per frame as a predictor. If clusters form that are qualitatively similar to the definitions of fixations, saccades, and smooth pursuit movements, the lower bounds of the distance per frame measure in each cluster can be used as a threshold that best identifies these activities in mouse movement data.

My future analyses will also include two uses of a video saliency model, DeepGaze MR (Tangemann et al., 2020). This model outputs predicted log density values for each x-y coordinate for each video frame, indicative of the likelihood of attention being directed to any particular content on a video frame. First, this model can be used to calculate the Information Gain metric (see Kümmerer et al., 2015) which here can be used to determine how well the model predicts the eye and mouse movement coordinates beyond a simple model only predicting central bias. I am also conducting an exploratory analysis by gathering the predicted log density values corresponding to each DIEM and MCBRD participant's x-y coordinates on each frame and using these predicted values in a larger model to determine the relative salience of each group's attended content over the course of different videos. Both of these methods can be used to determine if eye tracking and mouse tracking data from a blurred paradigm correspond to salient video content in a similar way. Second, I am using DeepGaze MR to analyze the relative salience of content in video frames after they have been filtered with different levels of blur. Analyses on these blurred videos can determine potential differences in the salience of the blurred periphery that may affect some videos differently depending on the original strength of the salient content in these videos. These analyses were originally considered for this validation study but have expanded enough that they are now a better fit for their own follow-up study.

A follow-up study is also warranted to examine the effect of computer type (laptop vs. desktop), browser, screen size, and mouse cursor speed on mouse movement patterns. Data on

these factors were collected in this study, but they could only be used for describing the general make-up of the study sample due to their relative sample sizes. A future study could experimentally manipulate these variables to determine their impact. For factors between participants that cannot be experimentally manipulated, like previous computer video game experience or wearing bifocals, a larger sample size is needed to increase the number of participants that respond from the uncommon categories. Future work will also be done on measuring the MCBRD paradigm refresh rate between different computers running different browsers and using different screen resolutions. These measures will be used in creating a comprehensive guide to inform researchers about what refresh rates they can expect when deploying a MCBRD paradigm on different device or browser types. Finally, a study with participants using a MCBRD paradigm while their eyes are tracked can be conducted to confirm that participants are attending to the high-resolution window instead of the blurred periphery.

Closing remarks

Through this study I have validated that a Mouse Contingent Bi-Resolution Display can be used as a measure of attention to Areas of Interest in videos and is best done so with a blur sigma of 0.45 dva, instead of a lower blur sigma of 0.15 dva or 0.30 dva. There are many potential use cases for a MCBRD paradigm in research. Because these paradigms are cheaper to implement than eye tracking and faster in terms of data collection (with the ability to run multiple subjects simultaneously online, instead of one at a time in a lab), mouse-contingent displays have the potential to become a huge asset for anyone studying attention to screen-based stimuli (e.g., consumer research, visual search research, video game research). With the methodology discussed here allowing us to run video stimuli in a mouse-contingent bi-resolution display, I similarly hope to extend this asset to the growing field of video-based attention studies.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1990). Eye-hand coordination: oculomotor control in rapid aimed limb movements. *Journal of experimental psychology: human perception and performance, 16*(2), 248.

Albert, R., Patney, A., Luebke, D., & Kim, J. (2017). Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP), 14*(4), 1–13.

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*(1), 1–12.

Anwyl-Irvine, A., Armstrong, T., & Dalmaijer, E. (2021). *MouseView. js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture*. *Behavior research methods*, 1–25.

Bates, D., Mächler, M., Bolker, B., & Walker, S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

Blackwell, A. F., Jansen, A. R., & Marriott, K. (2000). Restricted focus viewer: A tool for tracking visual attention. *International Conference on Theory and Application of Diagrams*, 162–177.

Bonikowski, L., Gruszczyński, D., & Matulewski, J. (2021). Open-source Software for Determining the Dynamic Areas of Interest for Eye Tracking Data Analysis. *Procedia Computer Science, 192*, 2568–2575.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80,* 1–28.

Burton, L., Albert, W., & Flynn, M. (2014). A comparison of the performance of webcam vs. Infrared eye tracking technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 1437–1441.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2018). What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(3), 740–757.

Cajar, A., Schneeweiß, P., Engbert, R., & Laubrock, J. (2016). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision, 16*(2), 1–19.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision research, 46*(26), 4333-4345.

Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, 281–282.

Collewijn, H., & Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision*, *8*(14), 20–20.

Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparitive Neurology*, *300*, 5–25.

Deng, J., Krause, J., & Fei-Fei, L. (2013). Fine-grained crowdsourcing for fine-grained recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of vision*, *10*(10), 28–28.

Dorr, M., Vig, E., & Barth, E. (2012). Eye movement prediction and variability on natural video data sets. *Visual cognition*, *20*(4–5), 495–514.

Drasdo, N., & Fowler, C. (1974). Non-linear projection of the retinal image in a wide-angle schematic eye. *The British Journal of Ophthalmology*, *58*(8), 709.

Duchowski, A. T., & Çöltekin, A. (2007). Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *3*(4), 1–18.

Duchowski, A. T., Cournia, N., & Murphy, H. (2004). Gaze-contingent displays: A review. *CyberPsychology & Behavior*, *7*(6), 621–634.

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science*, *15*(6), 431–431.

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1–20.

Findlay, J., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–721.

Geisler, W. S., & Perry, J. S. (1998, July). Real-time foveated multiresolution system for low-bandwidth video communication. In *Human vision and electronic imaging III* (Vol. 3299, pp. 294–305). International Society for Optics and Photonics.

George, A. (2019). Image based eye gaze tracking and its applications. *ArXiv Preprint ArXiv:1907.04325*.

Gorji, S., & Clark, J. J. (2018). Going from image to video saliency: Augmenting image salience with dynamic attentional push. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7501–7511).

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Guo, Q., & Agichtein, E. (2010). Towards predicting web searcher gaze position from mouse movements. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3601–3606).

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222.

Hoffman, D., Meraz, Z., & Turner, E. (2018). Limits of peripheral acuity and implications for VR system design. *Journal of the Society for Information Display*, *26*(8), 483–495.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795.

Holmqvist, K., & Andersson, R. (2017). Eye tracking: A comprehensive guide to methods. *Paradigms and Measures*.

Huang, J., White, R., & Buscher, G. (2012). User see, user point: Gaze and cursor alignment in web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1341–1350.

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications*, *2*(1), 1–30.

Jansen, A. R., Blackwell, A. F., & Marriott, K. (2003). A tool for tracking visual attention: The restricted focus viewer. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 57–69.

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). Salicon: Saliency in context. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1072–1080.

Jones, M. N., & Mewhort, D. J. K. (2004). Tracking attention with the focus-window technique: The information filter must be calibrated. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 270–276.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480.

Kaakinen, J. K., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *Quarterly Journal of Experimental Psychology*, *64*(7), 1372–1387.

Kim, J., Jeong, Y., Stengel, M., Akşit, K., Albert, R., Boudaoud, B., Greer, T., Kim, J., Lopes, W., & Majercik, Z. (2019). Foveated AR: Dynamically-foveated augmented reality display. *ACM Transactions on Graphics (TOG)*, *38*(4), 1–15.

Kim, N. W., Bylinskii, Z., Borkin, M. A., Gajos, K. Z., Oliva, A., Durand, F., & Pfister, H. (2017). BubbleView: An interface for crowdsourcing image importance maps and tracking visual attention. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *24*(5), 1–40.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916.

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye tracking for everyone. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2176–2184.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2015). Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences*, *112*(52), 16054–16059.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9(10):6*, 1–16.

Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, *13(12):11*, 1–20.

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka LeastSquares Means. R package version 1.3.3. https://CRAN.R-project. org/package=emmeans

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *45*, 251–266.

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS one*, *10*(11), e0142474.

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multi-resolutional display. *Journal of Experimental Psychology: Applied*, *8*(2), 99–117.

Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, *12*(6), 1057–1092.

Loschky, L. C., Ringer, R. V., Johnson, A. P., Larson, A. M., Neider, M., & Kramer, A. F. (2014). Blur detection is unaffected by cognitive load. *Visual Cognition*, *22*(3/4), 522–547.

Loschky, L. C., & Wolverton, G. S. (2007). How late can you update Gaze-contingent Multiresolutional Displays without detection? *Transactions on Multimedia Computing, Communications, and Applications*, *3*(4), 1–10.

Luebke, D., Hallen, B., Newfield, D., & Watson, B. (2000). *Perceptually Driven Simplification Using Gaze-Directed Rendering* (Technical Report CS-2000-04). University of Virginia, Department of Computer Science.

Lyudvichenko, V., & Vatolin, D. (2019). Predicting video saliency using crowdsourced mouse-tracking data. *ArXiv Preprint ArXiv:1907.00480*.

Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics-Physics Education Research*, *8*(1), 010122.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, *3*(1), 5–24.

Morys-Carter, W. L. (2021). *ScreenScale* (Computer software). Pavlovia.

Nielsen. (2020). The Nielsen total audience report: Advertising Across Today's Media. *Retrieved from https://www.nielsen.com/us/en/insights/report/2021/total-audience-advertising-across-todays-media/*

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342–360.

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*(2), 370–392.

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition, 20*(4–5), 457–494.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405.

Nyström, M., Hansen, D. W., Andersson, R., & Hooge, I. (2016). Why have microsaccades become larger? Investigating eye deformations and detection algorithms. *Vision research, 118*, 17–24.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 410–420.

O'Hare, L., & Hibbard, P. B. (2013). Visual discomfort and blur. *Journal of vision*, *13*(5), 7-7.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods, 51*, 195–203.

Peli, E., Yang, J., & Goldstein, R. B. (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America*, *8*(11), 1762–1774.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision research, 45*(18), 2397–2416.

Peterson, J. (2016). *The effect of blur on visual selective attention* (PhD Thesis). Kansas State University.

Peterson, J. J. (2018). *The interaction between visual resolution and task-relevance in guiding visual selective attention*. Kansas State University.

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, *29*(9), 1133–1151.

Quinn, N., Csincsik, L., Flynn, E., Curcio, C. A., Kiss, S., Sadda, S. R., Hogg, R., Peto, T., & Lengyel, I. (2018). The clinical relevance of visualising the peripheral retina. *Progress in Retinal and Eye Research*, *68*(January), 83–109.

Reingold, E. M., & Loschky, L. C. (2002a). Reduced saliency of peripheral targets in gaze-contingent multi-resolutional displays: Blended versus sharp boundary areas of interest. In A. T. Duchowski (Ed.), *Proceedings of the Eye Tracking Research & Applications Symposium 2002*. ACM.

Reingold, E. M., & Loschky, L. C. (2002b). Saliency of peripheral targets in gaze-contingent multiresolutional displays. *Behavior Research Methods, Instruments and Computers, 34*(4), 491–499.

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multi-resolutional displays: An integrative review. *Human Factors, 45*(2), 307–328.

Rodden, K., Fu, X., Aula, A., & Spiro, I. (2008). Eye-mouse coordination patterns on web search results pages. In *CHI'08 extended abstracts on Human factors in computing systems* (pp. 2997–3002).

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

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception, 18*(3), 347–361.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059–1074.

Tangemann, M., Kümmerer, M., Wallis, T. S., & Bethge, M. (2020, August). Measuring the importance of temporal features in video saliency. In *European Conference on Computer Vision* (pp. 667-684). Springer, Cham.

Thibos, L. N. (1998). Acuity perimetry and the sampling theory of visual resolution. *Optometry & Vision Science, 75*(6), 399–406.

van Diepen, P. M. J., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception, 27*(10), 1141–1151.

Võ, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of vision, 12*(13), 3–3.

Wade, N. J. (2010). Pioneers of eye movement research. *I-Perception, 1*(2), 33–68.

Wang, Z., Liu, Z., Li, G., Zhang, T., Xu, L., & Wang, J. (2021). Spatio-Temporal Self-Attention Network for Video Saliency Prediction. *arXiv preprint arXiv:2108.10696*.

Watson, A. B. (2014). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision, 14*(7), 15–15.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making, 16*(6), 1485–1505.

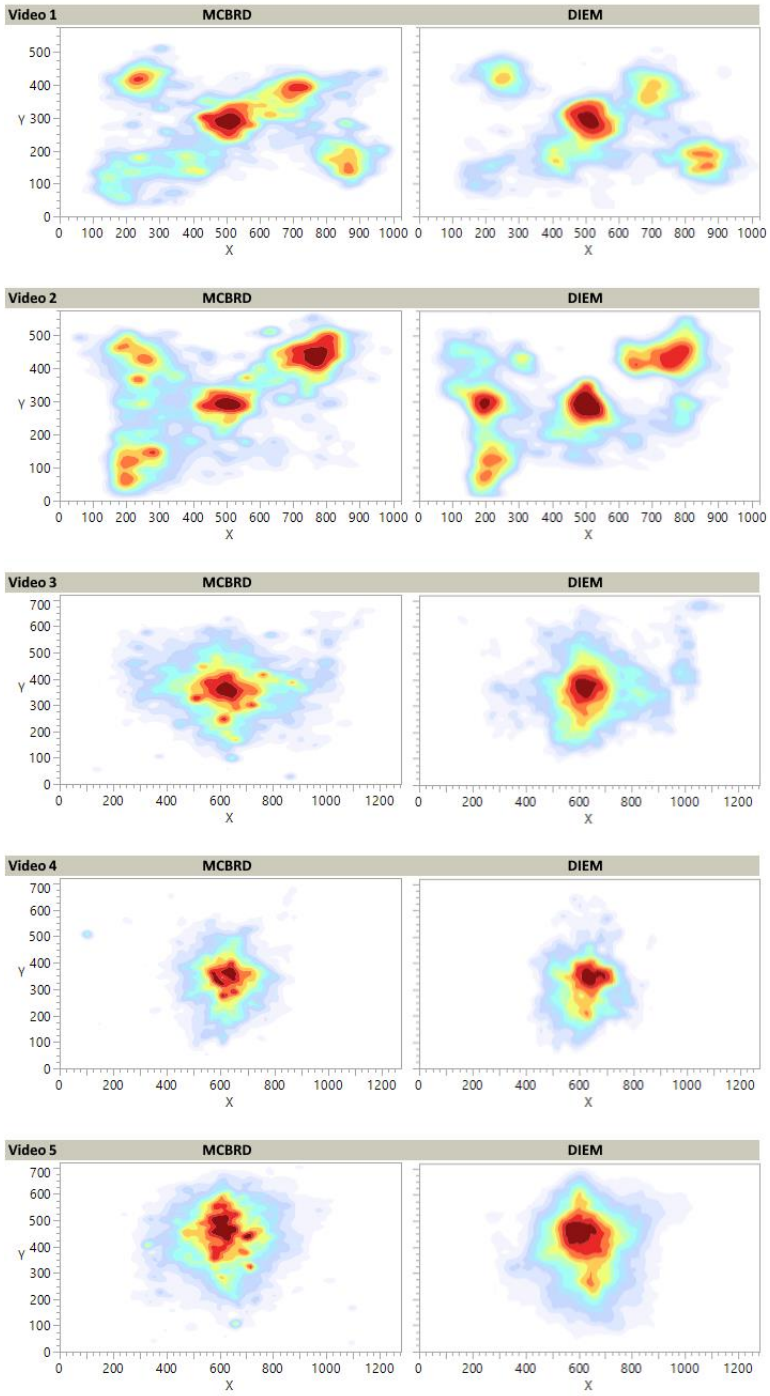
Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-based gaze estimation in the wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4511–4520

Appendix A - Heatmaps

This appendix contains Heatmaps generated from MCBRD x-y coordinate data (normalized to the scale of the original DIEM video presentation; left side) and from DIEM x-y coordinate data (right side) for all videos. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions. Figure axes are all scaled to the height and width of each video in pixels. As shown in the AOI analyses, participants viewing videos in MCBRD and eye tracking paradigms attended to similar regions of interest.

Figure 22

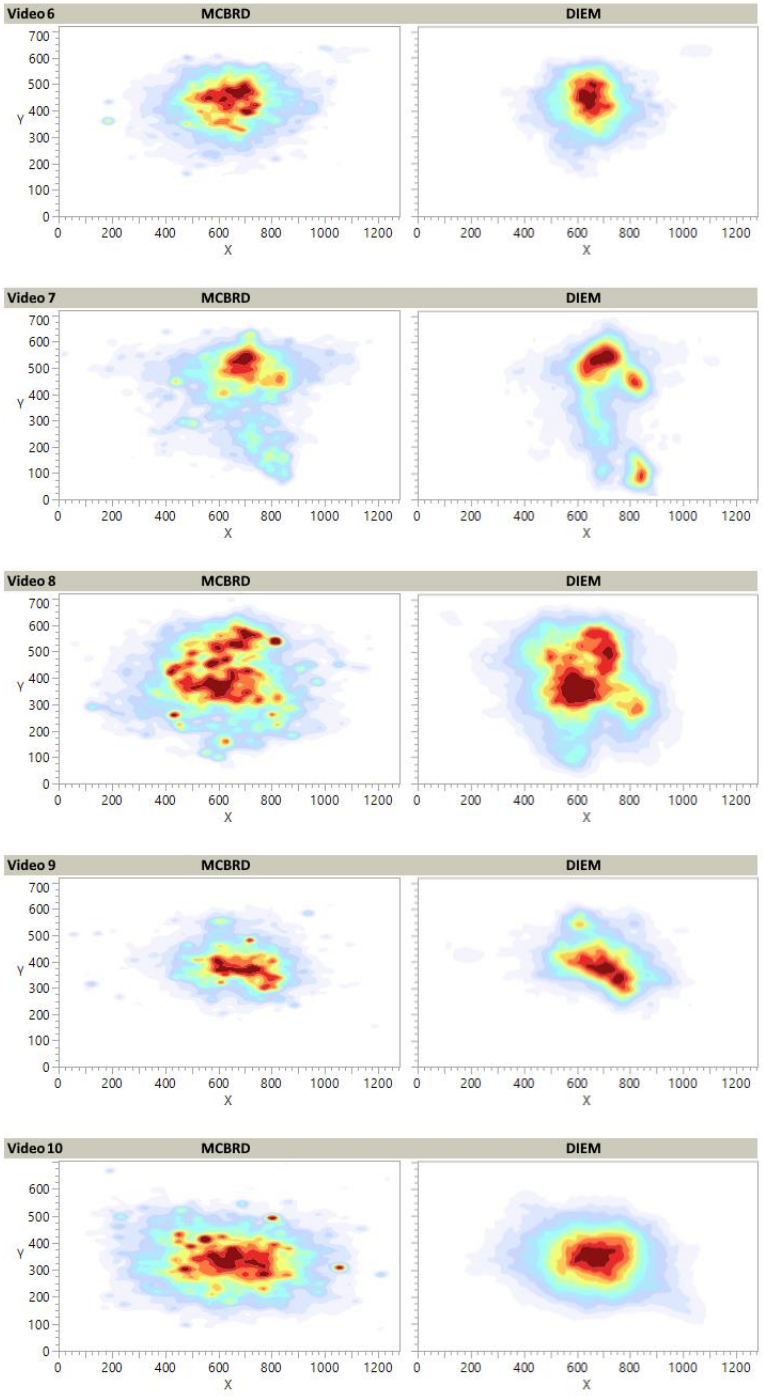
Heatmaps From MCBRD and DIEM Movements for Videos One Through Five



Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Figure 23

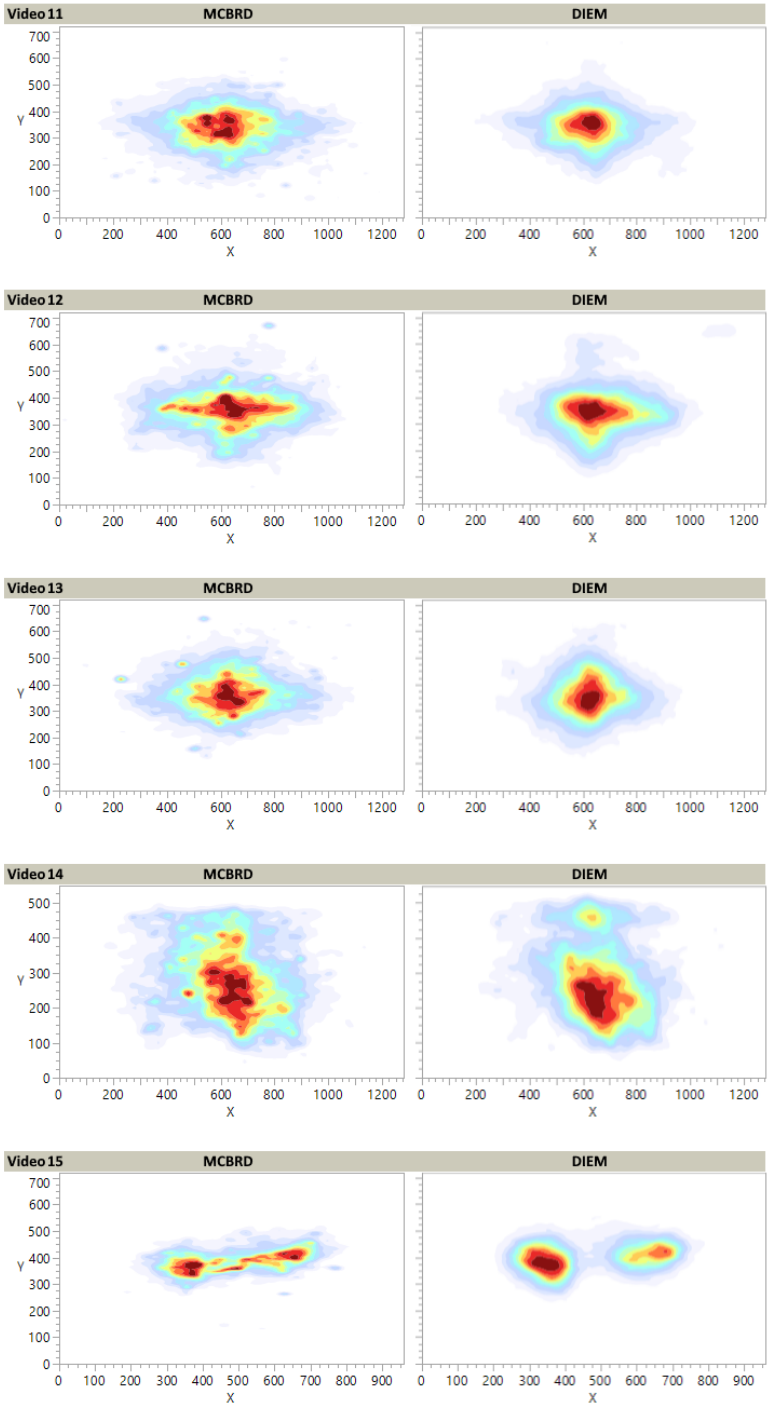
Heatmaps From MCBRD and DIEM Movements for Videos Six Through Ten



Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Figure 24

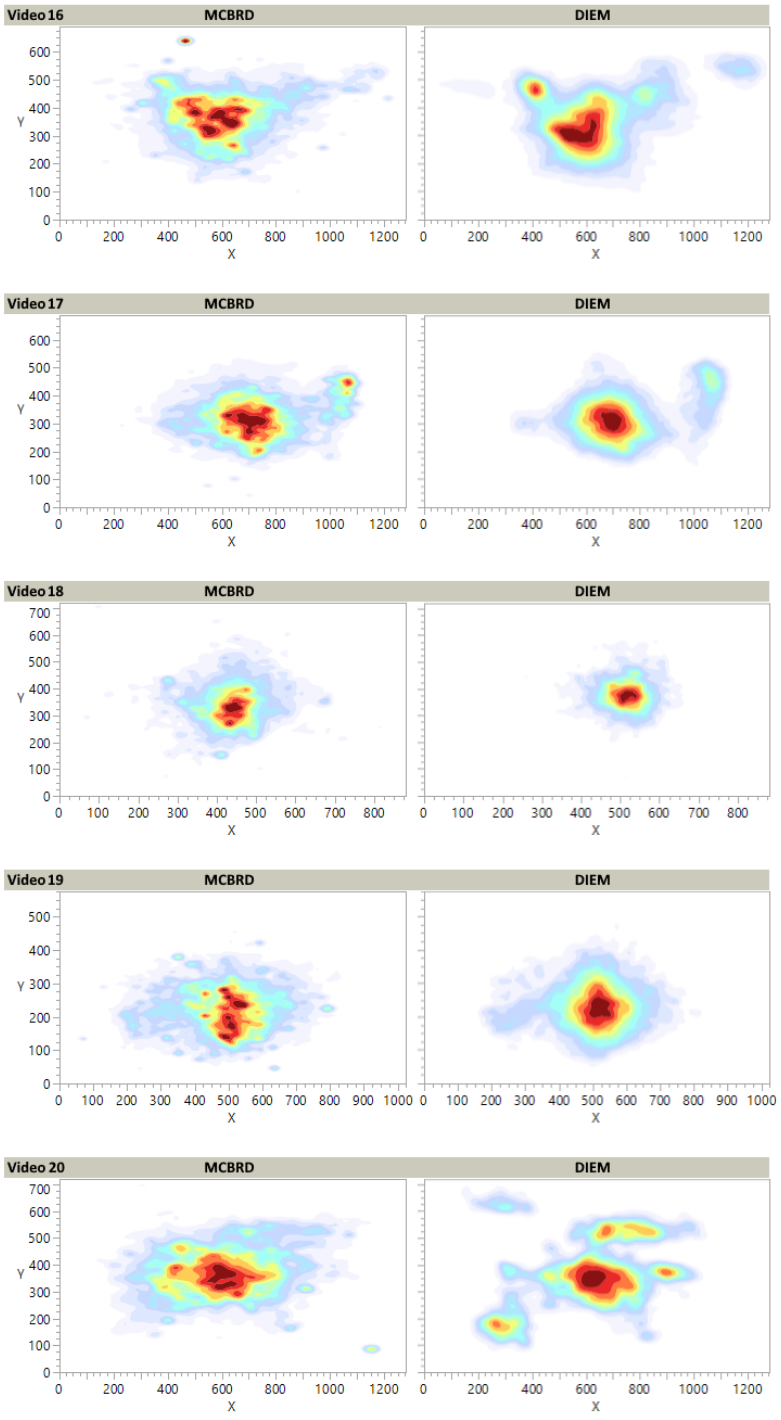
Heatmaps From MCBRD and DIEM Movements for Videos Eleven Through Fifteen



Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Figure 25

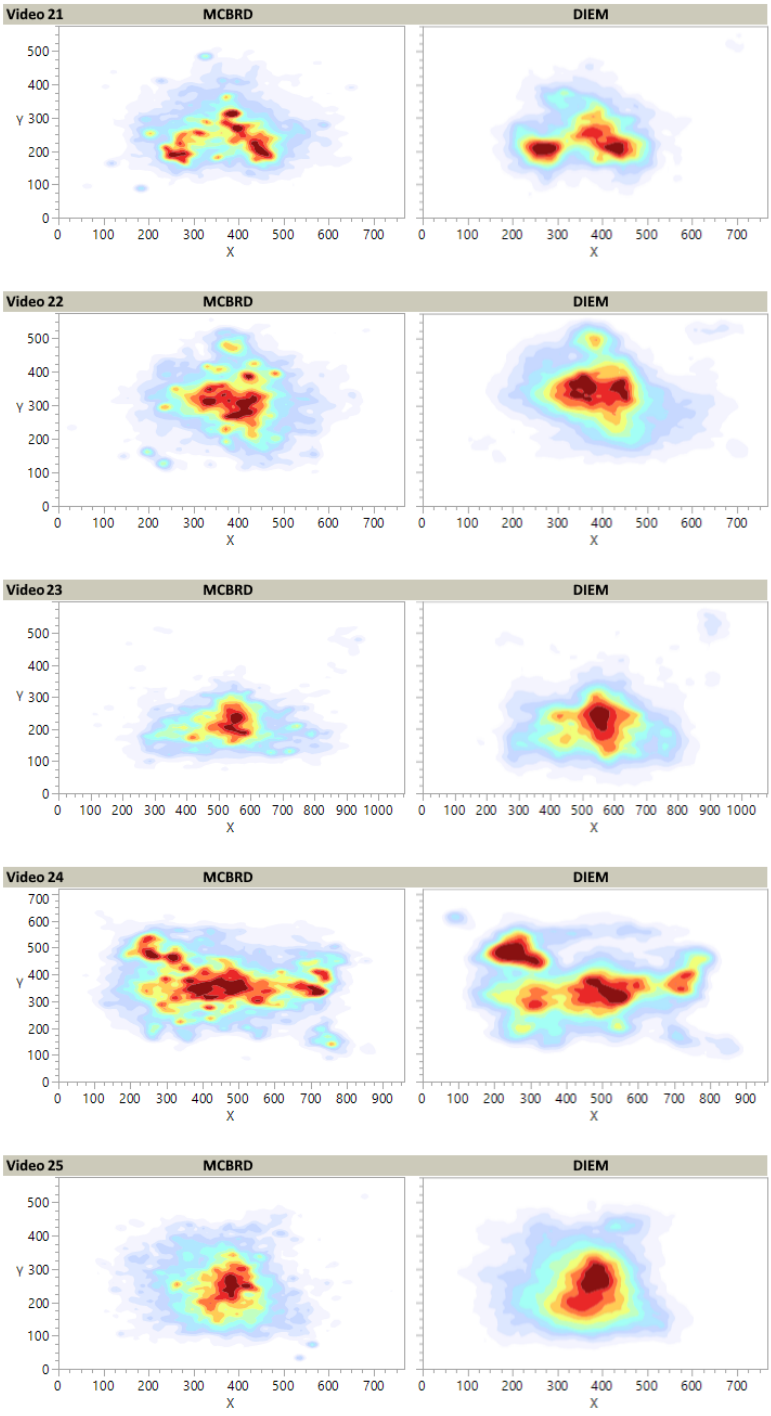
Heatmaps From MCBRD and DIEM Movements for Videos Sixteen Through Twenty



Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Figure 26

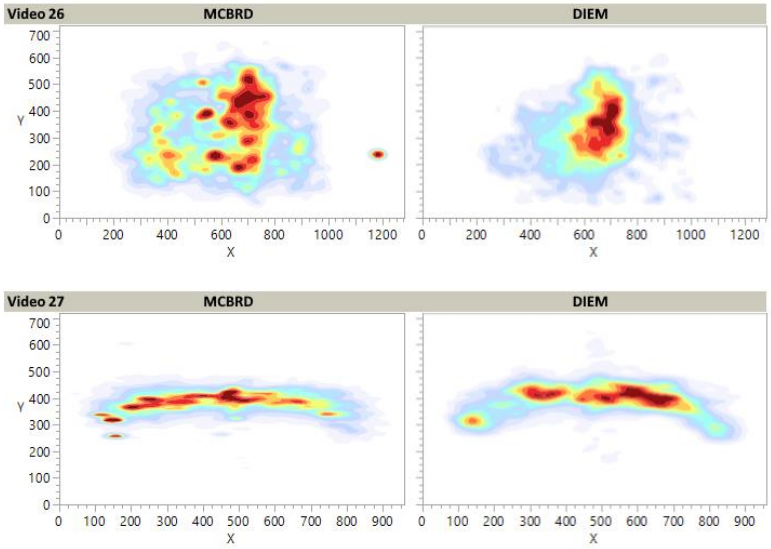
Heatmaps From MCBRD and DIEM Movements for Videos Twenty-One Through Twenty-Five



Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Figure 27

Heatmaps From MCBRD and DIEM Movements for Videos Twenty-Six Through Twenty-Seven



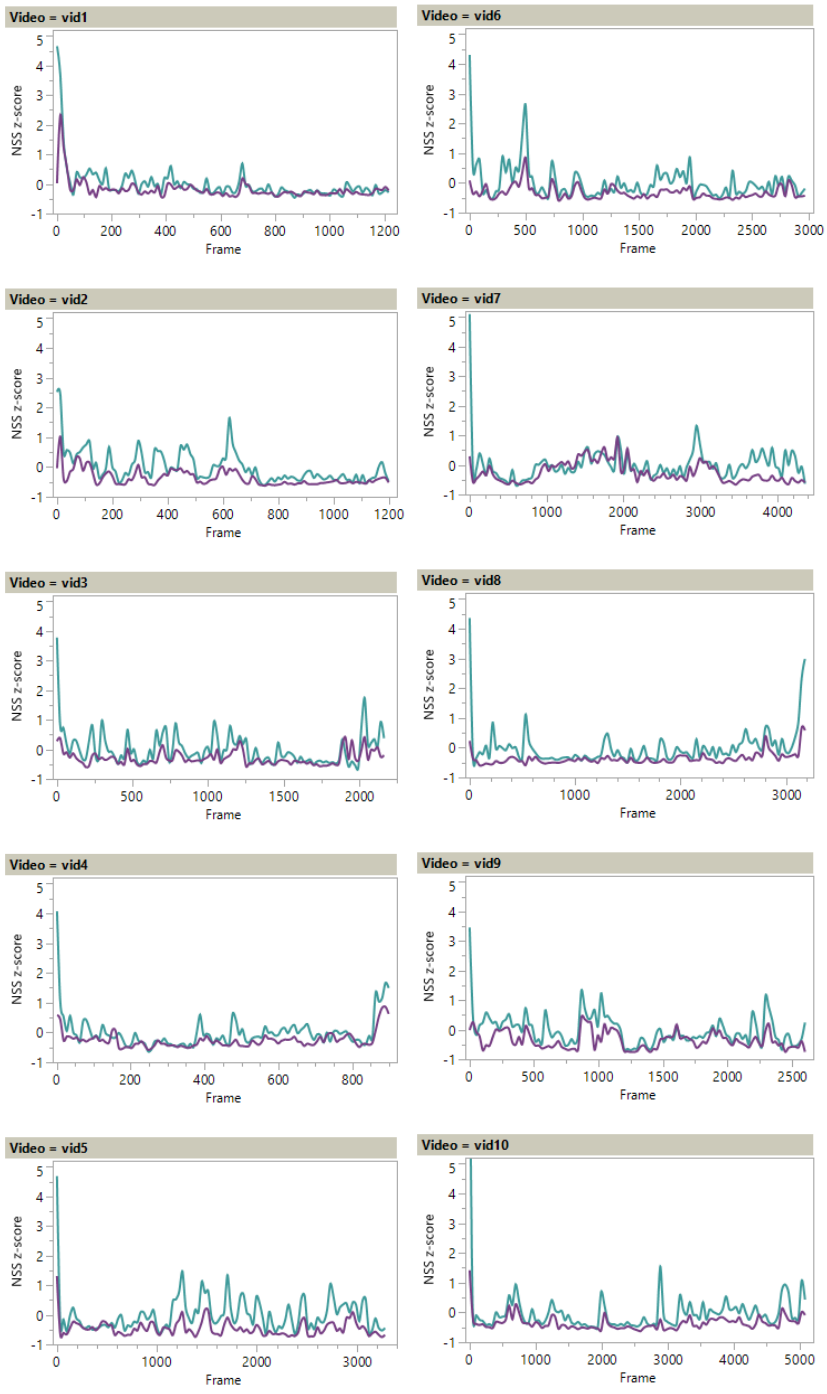
Note. MCBRD on left, DIEM on right. Heatmaps are created with 12 different levels, with colors closest to red representing the highest density of samples in those regions.

Appendix B - NSS Z-score by Frame Figures

This appendix contains figures depicting the mean z-score of gaze similarity of MCBRD (purple line in the figures) and DIEM (teal) participants compared to the DIEM reference gaze probability map for all videos. DIEM participants were compared to a probability map made from all DIEM data but theirs in a leave-one-out method. A positive z-score represents greater similarity than average, a z-score of zero represents average similarity, and a negative z-score represents lower similarity than average. Note: The videos for MCBRD participants were entirely blurred at the beginning, to prompt users to make the first mouse movement. Data recording for MCBRD participants began with the first mouse movement and was entered as NA values before that point. This resulted in lower NSS *z*-scores for MCBRD participants at the very beginning of each video.

Figure 28

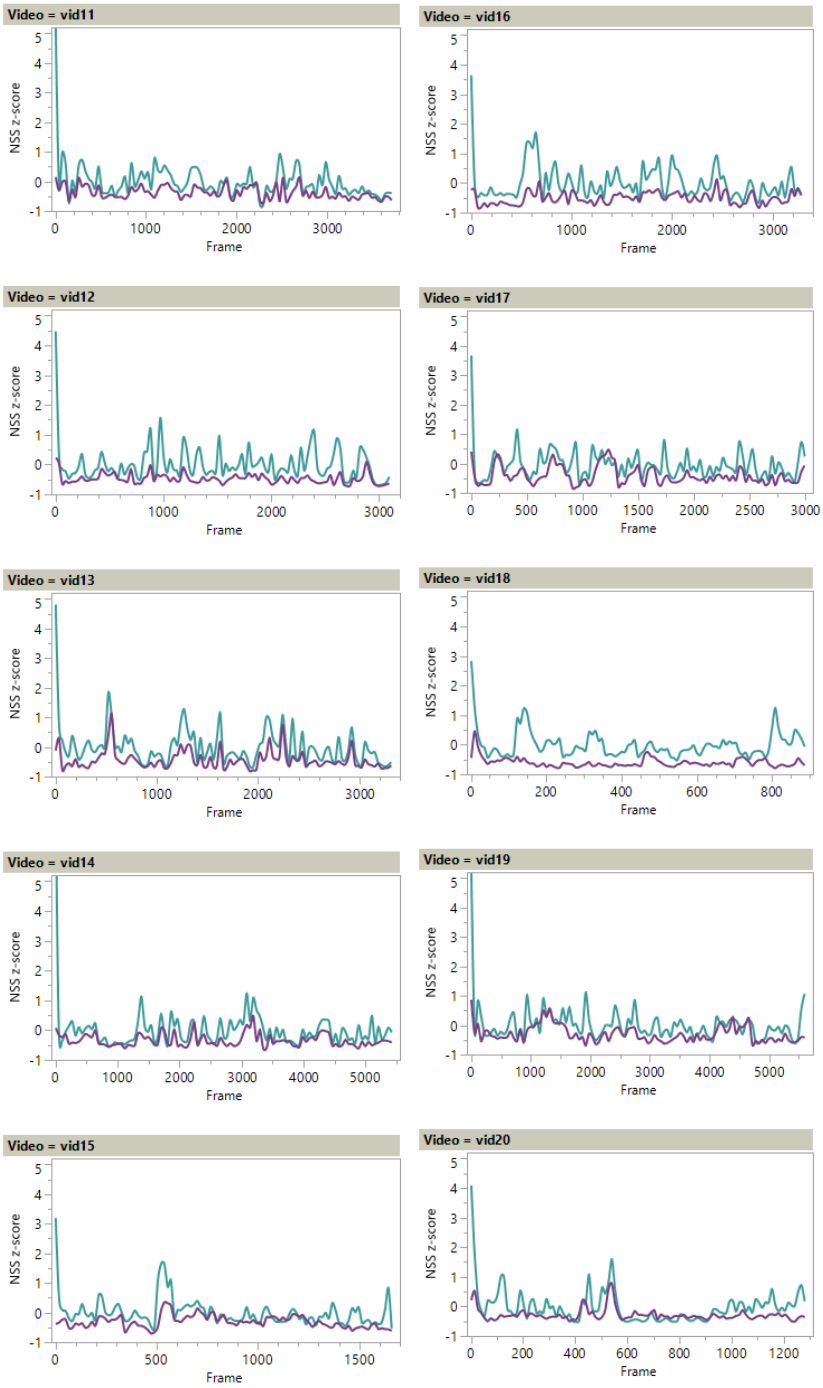
Gaze Similarity of MCBRD and DIEM Participants for Videos One Through Ten



Note. For videos 1-10, mean z-score of gaze similarity of MCBRD participants (in purple) and DIEM participants (in teal) compared to the DIEM reference gaze probability map.

Figure 29

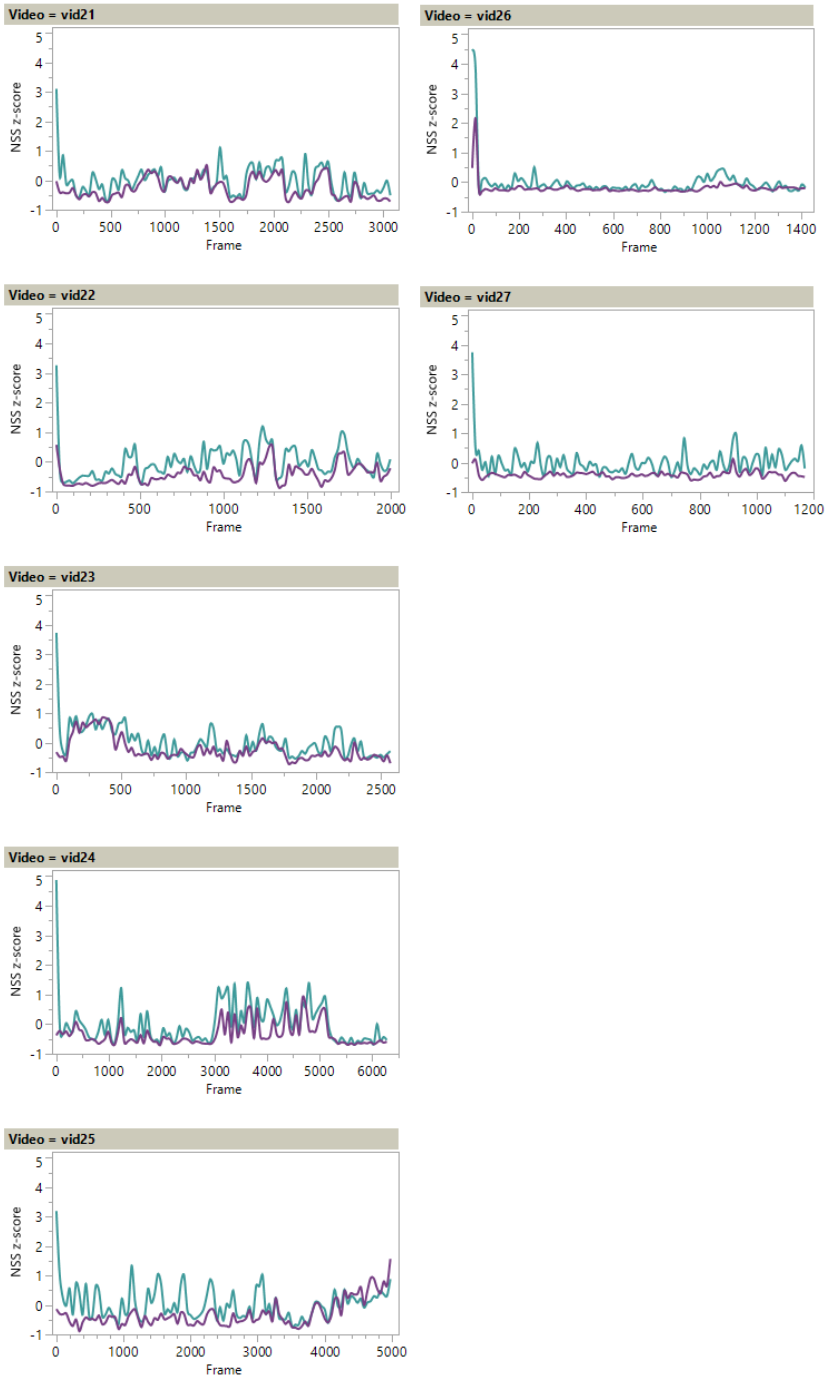
Gaze Similarity of MCBRD and DIEM Participants for Videos Eleven Through Twenty



Note. For videos 11-20, mean z-score of gaze similarity of MCBRD participants (in purple) and DIEM participants (in teal) compared to the DIEM reference gaze probability map.

Figure 30

Gaze Similarity of MCBRD and DIEM Participants for Videos Twenty-One to Twenty-Seven



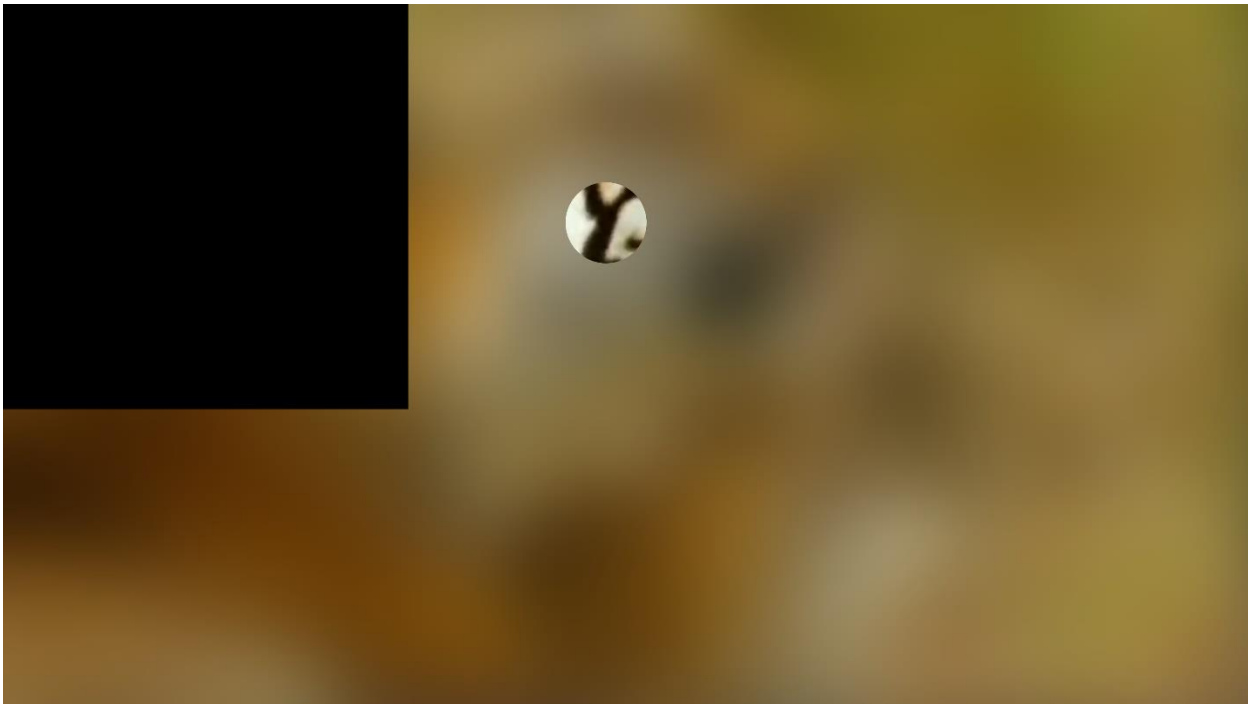
Note. For videos 21-27, mean z-score of gaze similarity of MCBRD participants (in purple) and DIEM participants (in teal) compared to the DIEM reference gaze probability map.

Appendix C - Refresh Rates

This appendix contains a still image from the refresh rate measurement program, and the refresh rate measurement results from all measured conditions. Refresh rates were gathered on three different computers and in three browsers per computer. Different mouse types and screen resolutions were also tested on one computer. This appendix also contains information about each tested computer to inform what refresh rates may be possible on similar devices. Refresh rates were all measured with a photodiode through a Chronos response box. Refresh rates were measured five times per condition, and the mean and standard deviations of these five measurements are featured in the tables below. Baseline refresh rates of all computer screens were 60 Hz. Slower refresh rates were found for Firefox across all computers (M = 32.6, SD = 16.1; see Table 14), for screen resolution above 2560x1600 (without Firefox, M = 31.9 , SD = 10.3; see Table 13), and in Safari in the one tested Mac computer (M = 11.9 , SD = 1.7; see Table 14). Average refresh rate across all 3 computers, using a screen resolution of 1440x900 or lower, running Chrome and Edge was M = 54.6 , SD = 8.9 (without the lower spec Lenovo computer's lower performance in Edge, $M = 58.7$, $SD = 3.2$.) Refresh rates were not changed by using a trackpad or mouse, see Table 15. The current recommendations are to run this paradigm in Chrome or Edge and at a screen resolution of 1440x900 if a refresh rate near 60 Hz is desired. In this study the videos were presented at a rate of 30 Hz, so values lower than 60 Hz are acceptable. Future refresh rate measurements will be added to create a comprehensive guide to inform researchers about what refresh rates they can expect from their sample.

Figure 31

A Still From the Refresh Rate Measurement Program



Note. The box in the top left corner would flicker between black and white whenever a mouse movement was detected and the blurred display updated. A blur sigma and window radius of 50px each were implemented for simplicity when quickly testing refresh rates in different browsers and on different computers.

Table 13

Mean Refresh Rates Across Screen Resolutions and Browsers

Browser	Resolution (px)	*M*	*SD*
Chrome	1440 × 900	60.25	1.00
Chrome	2560 × 1600	30.49	0.28
Chrome	3840 × 2160	24.40	0.84
Edge	1440 × 900	59.99	0.04
Edge	2560 × 1600	47.45	8.72
Edge	3840 × 2160	25.19	0.91
Firefox	1440 × 900	18.95	2.60
Firefox	2560 × 1600	28.88	0.20
Firefox	3840 × 2160	13.06	2.79

Note. Mean refresh rates collected from a Dell XPS laptop (specs in Table 16) across different screen resolutions and browsers. Screen resolutions of 1440 × 1900 px, 2560 × 1600 px, and 3840 × 2160 px were tested in Google Chrome, Microsoft Edge, and Mozilla Firefox browsers. Refresh rates were recorded five times per browser and resolution combination.

Table 14

Mean Refresh Rates From Three Computers Across Three Browsers

Computer	Spec Tier	OS	Browser	*M*	*SD*
Dell XPS	Mid	Windows 10	Chrome	60.25	1.00
Dell XPS	Mid	Windows 10	Edge	59.99	0.04
Dell XPS	Mid	Windows 10	Firefox	18.95	2.60
Lenovo IdeaPad	Low	Windows 10	Chrome	54.63	4.58
Lenovo IdeaPad	Low	Windows 10	Edge	38.59	5.75
Lenovo IdeaPad	Low	Windows 10	Firefox	54.05	4.37
MacBook Air	Mid	Mac OS Catalina	Chrome	59.76	0.33
MacBook Air	Mid	Mac OS Catalina	Firefox	24.87	1.72
MacBook Air	Mid	Mac OS Catalina	Safari	11.91	1.70

Note. Mean refresh rates collected from three different computers (specs in Table 16), testing three browsers each. The two Windows computers were tested in Google Chrome, Microsoft Edge, and Mozilla Firefox browsers. The Mac computer was tested in Google Chrome, Mozilla Firefox, and Apple Safari browsers. Dell XPS screen resolution was lowered to 1440x900 px to be closer to the other screen resolutions. Refresh rates were recorded five times per computer combination.

Table 15

Mean Refresh Rates Across Two Mouse Types, Trackpad and Corded

Resolution	Mouse Type	*M*	*SD*
3840 × 2160	Trackpad	24.40	0.84
3840 × 2160	Corded	28.12	1.14
3840 × 2160	Wireless	26.60	2.20
1440 × 900	Trackpad	60.25	1.00
1440 × 900	Corded	60.00	0.75
1440 × 900	Wireless	59.99	0.83

Note. Mean refresh rates collected from a Dell XPS laptop (specs in Table 16) running Google Chrome, tested using 1440 × 900 px and 3840 × 2160 px screen resolutions and two mouse types, trackpad and corded. Refresh rates were recorded five times per resolution and mouse type combination.

Table 16

Technical Specifications for All Computers Used for Refresh Rate Testing

Computer	OS	CPU	UB	PM	RAM	Screen Resolution	Spec Tier
Dell XPS 15 9550	Windows	Intel Core i7-7700HQ	62nd	6953	16 GB DDR4	3840 × 2160	Mid
Lenovo IdeaPad 110	Windows	AMD A6-7310 APU	33rd	1736	4 GB DDR3	1366 × 768	Low
MacBook Air 13 2017	Mac OS	Intel Core i7-5650U	56th	3044	8 GB DDR3	1440 × 900	Mid

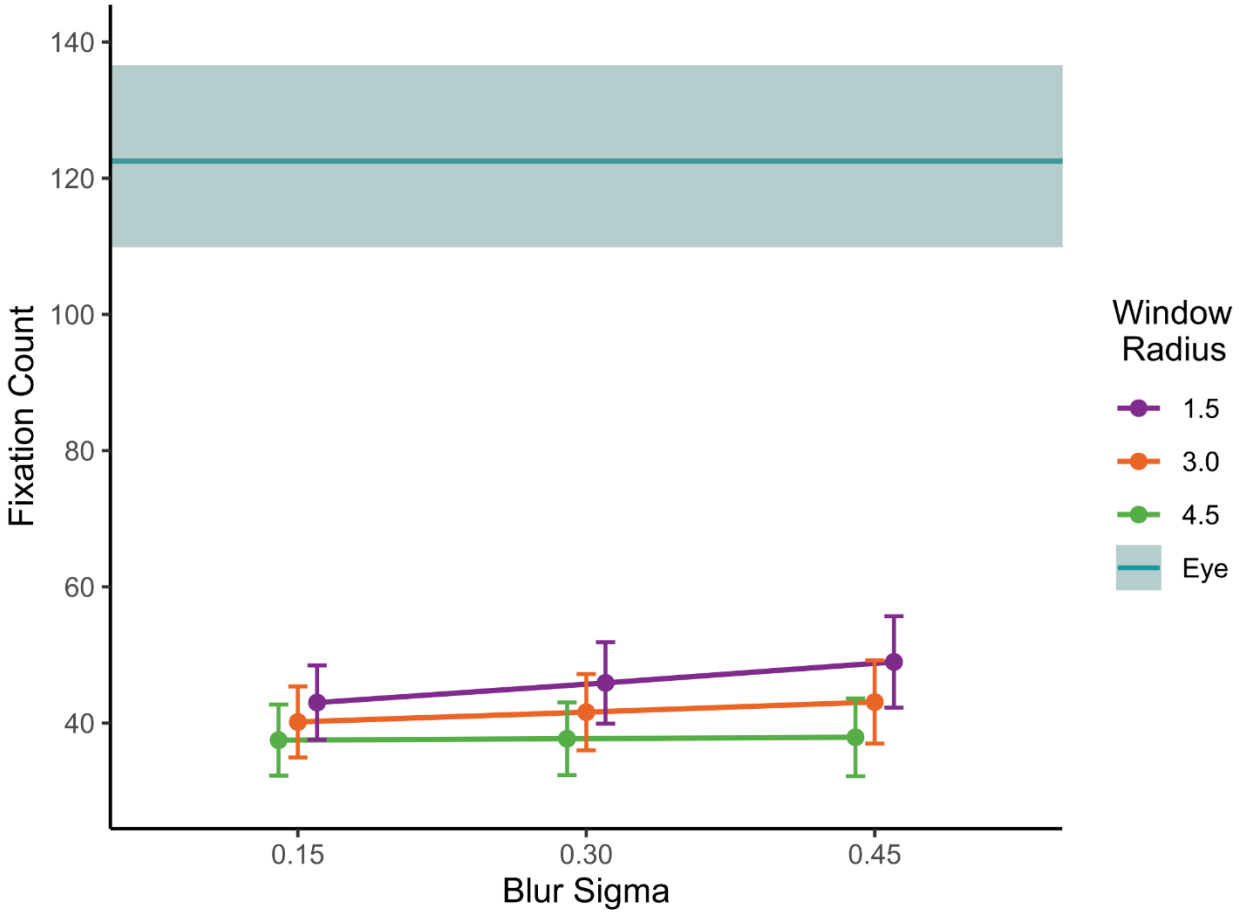
Note. Benchmark ratings from User Benchmark (UB; https://cpu.userbenchmark.com/; higher percentiles are better)and Passmark CPU Benchmark (PM; https://www.cpubenchmark.net/; scores are relative to each other, higher scores are better) are provided to compare CPU power across brands. For RAM, more RAM is better, and DDR4 is newer, faster RAM than DDR3. A subjective rating of Spec Tier is added for easier interpretation of these values, in comparison to other computers one could expect in a subject population.

Appendix D - Trial Fixation Analyses

This appendix includes figures for trial fixation count and fixation duration analyses that were run for data visualization purposes. These fixations were classified as any time period where the participant's measured x-y coordinates move no more than 1 dva, for a minimum duration of 100 ms (common thresholds, see Holmqvist & Andersson, 2017). These figures show the differences in movement types between eye and mouse movement data. These differences suggest it may be worthwhile to determine new thresholds to describe the different motions made by mouse movements.

Figure 32

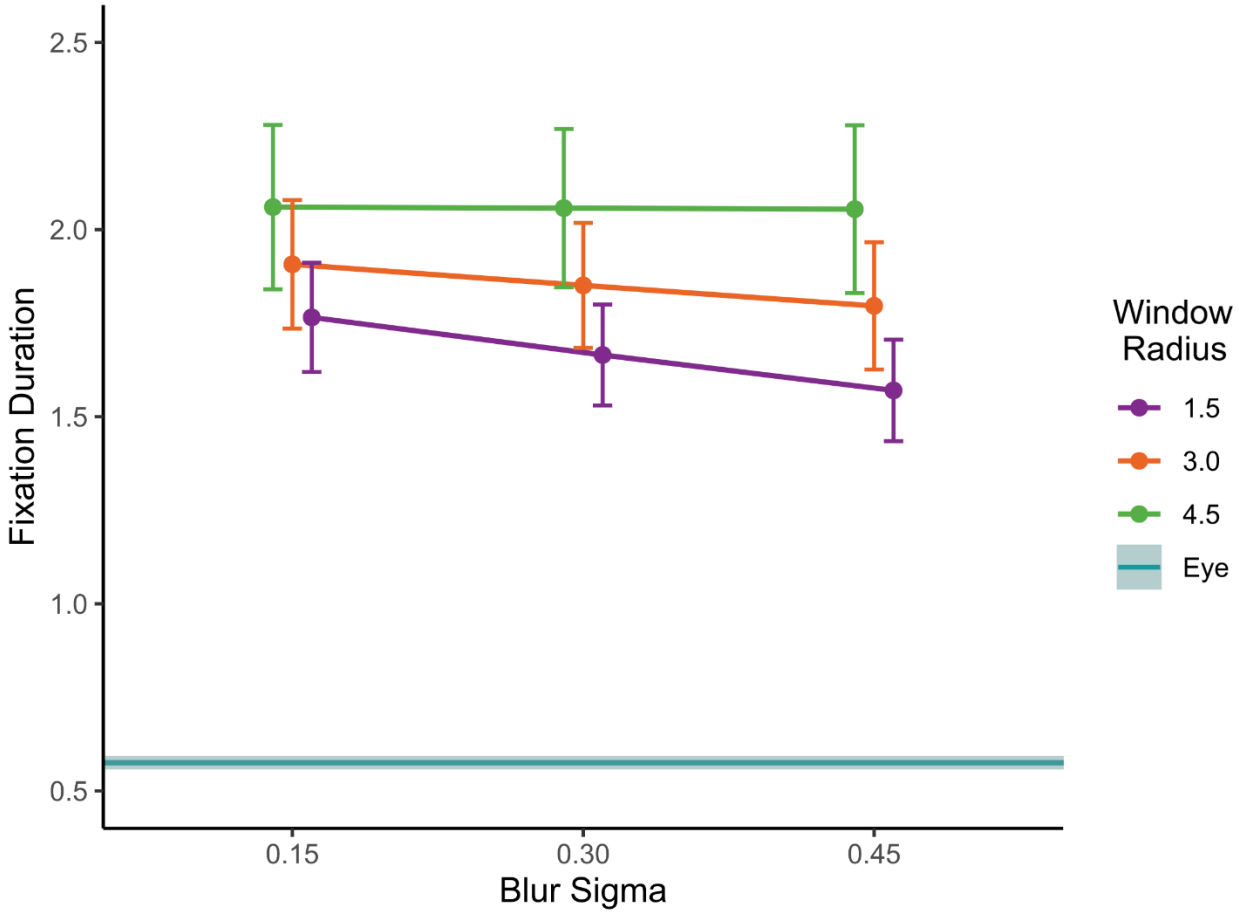
Estimated Marginal Means From Poisson Regression Predicting Number of Fixations



Note. Estimated marginal means from a multilevel Poisson regression with blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction predicting number of fixations. Random intercept effects and slope effects of blur, window, and their interaction for both video and participant were included in this analysis. Error bars are equal to 1 *SE*. All values are back transformed from the log scale of the original model. DIEM mean (122.5) is featured as a solid teal line with an error ribbon representing 1 *SE* (asymmetric *SE* with a mean of 13.4). For information on how this baseline was generated, see DIEM Baselines above.

Figure 33

Estimated Marginal Means From Gamma Regression Predicting Fixation Duration



Note. Estimated marginal means from a multilevel gamma regression with blur sigma (continuous predictor, centered by subtracting the mean of 0.3), window radius (continuous predictor, centered by subtracting the mean of 3), and their interaction predicting fixation duration (in seconds). Random intercept effects and slope effects of blur, window, and their interaction for both video and participant were included in this analysis. Error bars are equal to 1 *SE*. All values are back transformed from the log scale of the original model. DIEM mean (0.57) is featured as a solid teal line with an error ribbon representing 1 *SE* (asymmetric *SE* with a mean of 0.02). For information on how this baseline was generated, see DIEM Baselines above.

Appendix E - DIEM Video Names

This appendix includes a table with the names of all included videos as they are listed in the DIEM dataset (Mital et al., 2011; see https://thediemproject.wordpress.com/).

Table 17

Names of Video Stimuli Used in This Project From the DIEM Dataset

Video	Name
1	advert_bbc4_bees_1024x576
2	advert_bbc4_library_1024x576
3	advert_bravia_paint_1280x720
4	advert_iphone_1272x720
5	documentary_adrenaline_rush_1280x720
6	documentary_coral_reef_adventure_1280x720
7	documentary_discoverers_1280x720
8	documentary_dolphins_1280x720
9	documentary_mystery_nile_1280x720
10	documentary_planet_earth_1280x704
11	game_trailer_bullet_witch_1280x720
12	game_trailer_ghostbusters_1280x720
13	game_trailer_lego_indiana_jones_1280x720
14	game_trailer_wrath_lich_king_shortened_subtitles_1280x548
15	home_movie_Charlie_bit_my_finger_again_960x720
16	movie_trailer_ice_age_3_1280x690
17	movie_trailer_quantum_of_solace_1280x688
18	music_gummybear_880x720
19	music_red_hot_chili_peppers_shortened_1024x576
20	music_trailer_nine_inch_nails_1280x720
21	news_bee_parasites_768x576
22	news_sherry_drinking_mice_768x576
23	news_us_election_debate_1080x600
24	news_video_republic_960x720
25	news_wimbledon_macenroe_shortened_768x576
26	university_forum_construction_ionic_1280x720
27	pingpong_closeup_rallys_960x720
28	BBC_wildlife_special_tiger_1276x720
29	sport_golf_fade_a_driver_1280x720

Note. Names of all video stimuli used in this project from the DIEM dataset (Mital et al., 2011).Videos 28 and 29 were practice videos shown at the beginning of the experiment.