Dynamics of spatial attention during motion tracking: Characterization and modeling as a function of motion predictability

by

Abdullah Zafar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Kinesiology

Waterloo, Ontario, Canada, 2023

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Efficient information processing in ecological environments relies on spatial attention to selectively process relevant areas in the visual field. Attention has been shown to be biased ahead of simple, uniform target motion during smooth pursuit [Khan et al., 2010]. However, real-world motion varies in predictability, and as such this study aimed to: a) determine how motion predictability affects attentional bias, b) characterize how visual attention adapts to changes in motion predictability, and c) implement a computational model of visual attention during motion tracking.

Ten high-performance team sport athletes (5 male, 5 female) and ten healthy, young adults (5 male, 5 female) visually tracked a target moving at varying predictability levels. A probe was flashed ahead or behind target motion (2° or 6°), and manual response times (MRT) to probes were collected to indicate attention level. To investigate the temporal dynamics of attentional bias, a second tracking task was performed where the target changed predictability levels mid-trial. The effects of group, motion predictability, and probe distance, time & location on MRT bias were examined. Finally, a state-space model (input: target motion, output: attentional bias) was trained and tested on the motion tracking and MRT data using a 5-fold cross-validation.

MRT were significantly biased in athletes (distance=2°) and adults (distance=2°,6°) during predictable motion ($p<0.01$). There was no MRT bias for semi- and un-predictable motions. Furthermore, MRT bias took longer to accumulate, than it did to de-accumulate ($p<0.01$). Eye movements showed that catch-up saccades were larger ($p<0.01$) and more frequent ($p<0.01$) during unpredictable motion phases, and gradually reduced in size and frequency during sustained predictable motion. Cross-validation results demonstrated that the state-space model performance in predicting attentional bias had a mean absolute error of 18.6% (SD=0.04%).

In conclusion, the distribution of spatial attention during motion tracking is dependant on motion predictability, and the accumulation of bias ahead of target motion takes longer than de-accumulation. These results indicate a conservative attentional allocation scheme that introduces bias based on predicted future errors in motion extrapolation. The state-

space model developed based on these experimental results may extend existing dynamic saliency frameworks to factor in the effects of motion tracking on spatial attention.

Acknowledgements

بِسْمِ اللهِ الرَحمٰنِ الرَحيْمِ

In the name of God, the most Gracious, the most Merciful

First and foremost, all praise is due to God for the ability to complete the writing of this thesis.

I would like to express my gratitude to my supervisor, Dr. Ewa Niechwiej-Szwedo for the opportunity to learn and work under her in the Developmental Visuo-motor Neuroscience Lab. Thank you to Dr. Michael Barnett-Cowan and Dr. Kristine Dalton for their consideration and guidance as members of my thesis committee.

I would also like to acknowledge my family and friends who were a consistent means of support and encouragement. To my mother, who is always motivating me to excel in any work I pursue; to Huma and Meraj, for their openness in discussing ideas and providing feedback; to my wife, Elizabeth, whom I had the pleasure to marry over the course of this work and who has been a vital source of tranquility; truly, thank you all.

Table of contents

List of figures

List of tables

Chapter I: Review of the literature

Introduction

The natural visual scene can be considered a complex system of dynamic stimuli, composed of several objects at various locations, of varying shape, and following a variety of movement patterns. Such scenes present themselves abundantly in real life: from the pedestrian crosswalk, with vehicles, bicycles and pedestrians all needing ample consideration before navigating the road, to athletic scenes where the movements of the ball, opposition and teammates are critical pieces of information in planning actions. In order to efficiently acquire useful scene information, a selective process of visual attention is employed to maximize visual processing of relevant scene features which may facilitate future scene prediction. For example, while visually tracking the motion of a ball in flight, the allocation of visual attention to regions ahead of ball motion may facilitate the extrapolation of the ball's trajectory, and by extension the future location of the ball.

Smooth pursuit eye movements occur when the eye rotates while fixing a moving target on the fovea [Martin, 2012]. These eye movements are key for tracking objects in real life scenes, and a number of studies have shown that pursuit, rather than fixation, improves motion prediction ability [Spering et al., 2011, Bennett et al., 2010]. Additionally, several studies have also examined the spatial allocation of visual attention during smooth pursuit using visual discrimination tasks [Lovejoy et al., 2009], visual reaction time tasks [Khan et al., 2010, Donkelaar and Drew, 2002], or EEG tagging [Chen et al., 2017]. While visual discrimination tasks demonstrated no bias in attention allocation, the reaction time tasks and EEG tagging demonstrated that the allocation of attention is biased ahead of pursuit target motion. The attentional bias was shown to be modulated by speed of pursuit, with a faster moving target resulting in more attention being biased ahead of the target [Donkelaar and Drew, 2002].

To date, the role of attentional bias during smooth pursuit has only been hypothesized as a function of motion extrapolation mechanisms to facilitate visual pre-processing of areas where the system predicts the target to be [Khan et al., 2010]. However, mo-

tion prediction occurs on several scales of time. Simple, uniform motion tracking requires time to visually process motion and send motor commands, meaning some level of prediction mechanism is required to not be constantly lagging behind the target [Brouwer et al., 2002]. However, longer range motion dynamics (i.e. motion patterns observed over longer time-frames) also interact with motion extrapolation mechanisms, as shown with periodic motion [Verfaillie and d'Ydewalle, 1991]. Therefore, the temporal dynamics of visual spatial bias during smooth pursuit should be further investigated to determine whether the bias arises due to local motion characteristics or rather emerges due to longer range motion history.

The visual system

The visual system comprises a variety of pathways in order to effectively process the wide range of visual information sensed by the retinas. Generally, the visual system can be separated into a dorsal stream and a ventral stream [Goodale and Milner, 1992]. The dorsal stream, involving the posterior parietal cortex, is primarily involved in localizing objects and processing motion, while the ventral stream, involving the inferior temporal lobe, is primarily involved in object and facial recognition. The focus of the following review of the visual system will be the relevant structures and pathways which facilitate visual motion processing, as well as those which produce smooth pursuit eye movements for visual motion tracking.

Visual motion processing

The segregation of functional neural pathways occurs almost immediately as visual signals from the photoreceptors of the retina project onto the lateral geniculate nucleus (LGN) via the optic nerve [Kandel et al., 2000]. Within the LGN, small parvocellular (P) cells form P-layers and larger magnocellular (M) cells form M-layers. The P cells demonstrate high spatial acuity, offset by low temporal acuity, and are also sensitive to color. By contrast, the M cells demonstrate high temporal acuity, offset by low spatial acuity, and are insensitive to color [Kandel et al., 2000]. The distinction of these two types of visual

content continues along the visual processing pathway as signals project from the LGN onto the primary visual cortex (V1) via optic radiations [Nealey and Maunsell, 1994].

The primary visual cortex (or striate cortex) consists of neurons which are mapped retinotopically and have small receptive field size. In macaque monkeys, the receptive field sizes in V1 range from <1 degree of visual angle in the fovea to about 5 degrees of visual angle at 40 degrees of eccentricity to the fovea; by contrast, receptive field sizes of neurons in the secondary visual cortex (V2) are up to 20 degrees of visual angle in size at similar eccentricities [Gattass et al., 1981]. As such, the retinotopic spatial encoding of objects in V1 is highly accurate and matches the subjective visual field well. The magnocellular and parvocellular projections from the LGN arrive at separate locations in V1: the M-cells project onto layer 4B/4Ca while the P-cells project onto layer 4A/4Cb [Martin, 2012]. There is also a columnar organization to the striate cortex, where "orientation" columns have been observed as groups of neurons which all respond to the same orientation stimulus and span across multiple cortical layers [Hubel and Wiesel, 1974]. In particular, layer 4B contains the highest concentration of cells which are sensitive to direction and provides the largest output from V1 to the middle temporal region [Essen and Maunsell, 1983].

The middle temporal (MT) or area V5 has been observed to contain neurons sensitive to motion direction [Zeki, 1980]. A variety of types of directional sensitivity exist, inferred from the level of neuronal activation when presented with different motion stimuli; for example, certain columns of neurons respond maximally when opposing motion is sensed in each eye (i.e. detection of movement towards or away from the observer) whereas other columns respond only to centripetal motion, and other still to select motion directions [Zeki, 1974]. The response fields of neurons can vary in size but are generally larger than those in V1, demonstrating a capacity for visual motion integration [Essen and Maunsell, 1983]. In particular, an "aperture" problem arises when sampling motion information from neurons with small response fields - namely, the limit of the response field size limits the amount of motion that can be sampled. To overcome such limitations, motion detected from several neurons with smaller receptive fields are inte-

grated by neurons with larger receptive fields to generate a larger motion image, as is the relation with V1 and MT [Pack, 2001]. Finally, inputs to MT are not limited by the magnocellular dominant afferents of V1, but evidence from trans-synaptic tracing and EEG studies have demonstrated that direct input from parvocellular pathways of the LGN exist, thereby providing MT with a robust set of chromatic, spatial, temporal signals for processing visual motion [Nassi et al., 2006, Schoenfeld et al., 2002].

Smooth pursuit eye movement system

Smooth pursuit eye movements are slow, conjugate eye movements used to continuously track a target on the fovea [Martin, 2012]. In order to smoothly track a visual target, contributions from various structures are required, from motion processing to detect velocities [Ilg, 2008] to motion prediction for overcoming sensorimotor delays [Barnes and Asselman, 1991]. The focus of the following review of the smooth pursuit system will be to highlight key structures which interact with processing, prediction, and attention functional networks.

The middle temporal (MT) and medial superior temporal (MST) lobes are heavily involved in motion processing due to the higher number of neurons selective for directional motion [Essen and Maunsell, 1983] and speed [Duffy and Wurtz, 1997]. Neurons selectively tuned for speed in these regions guide the initial velocity estimates needed for pursuit eye acceleration, rather than using the rate of neuronal activation across adjacent receptive fields [Priebe et al., 2001]. The effect of lesions in MT and MST on smooth pursuit have been noted, resulting in deficits in the ability to match target speeds during pursuit initiation, as well as deficits in the ability to maintain eye velocity during smooth pursuit [Newsome et al., 1985, Dursteler and Wurtz, 1988]. Furthermore, MT and MST can be affected by attentional modulation during smooth pursuit [Recanzone and Wurtz, 2000]. When neurons in these regions are presented with conflicting motion in the same receptive field, perception of that motion can either be an averaged motion vector or a biased motion vector. Given a target to overtly attend and pursue that conflicts with motion of distractor stimuli in the same receptive field, the perception of averaged motion is

observed up to about 450 milliseconds (ms), after which the perceived motion response becomes biased towards the attended target [Recanzone and Wurtz, 2000].

Neurons in the frontal eye fields (FEF) have been shown to have directionally selective responses during smooth pursuit which are broadly tuned [Gottlieb et al., 1994]. The firing rate of these neurons which discriminate for motion direction leads pursuit eye movement onset by an average of about 8 ms, with some neurons leading onset by about 25 ms [Tanaka and Lisberger, 2002]. However, the inhibition of FEF neurons has a more pronounced effect on the initial pursuit acceleration and causes a reduction steady-state velocity, rather than on pursuit initiation latency [Shi et al., 1998]. Conversely, electrical stimulation of the FEF increases pursuit velocity gain irrespective of motion direction [Tanaka and Lisberger, 2001]. Altogether with parallel results indicating the contribution of FEF neurons in selective visual spatial attention [Schall et al., 1995, Beauchamp et al., 2001], the omni-directional sensitivity of the FEF to motion and resultant eye movement speed may imply its role in applying target selection for the generation of oculo-motor plans [Tanaka and Lisberger, 2001].

Target selection for eye movements has also been observed in the superior colliculus (SC) which has retinotopically arranged superficial layers that encode for covert visual attention [Carello and Krauzlis, 2004, Katyal et al., 2010]. The intermediate and deep layers of the superior colliculus also contain "fixation" neurons which are active when targets are close to the fovea and suppressed when the fovea-target position is large, in order to produce target-directed saccades [Krauzlis et al., 2000]. However, activity in these neurons is also present during smooth pursuit eye movements where the fovea-target distance is relatively small. As such, the role of these "fixation" neurons in the SC is to encode error signals between the fovea and target in order to determine the generation of fixation, smooth pursuit, or saccades. As error increases, the suppression of these neurons allows for catch-up saccades to be produced, while as the error decreases, the activation of these neurons can result in either fixation or pursuit [Krauzlis et al., 2000].

Visual spatial attention

Visual spatial attention refers to the selective processing of areas in the visual field. The need for selectivity is due to a limit on neural processing capabilities, arising from a finite metabolic supply to the brain and the constant metabolic cost of activating neurons [Clarke and Sokoloff, 1999, Lennie, 2003]. As such, selectivity allows for a greater amount of resources to be allocated to processing a smaller portion of the visual field, rather than spreading resources across larger areas [Müller et al., 2003]. Several theories of spatial attention have been proposed, along with studies investigating the neural control mechanisms involved in attention allocation; however, investigations of the spatial allocation of attention during visual motion tracking have been limited to uniform motion warranting further investigation of complex motion.

Theories of visual spatial attention

Evidently selective processes of attention must exist from both the physiological standpoint, that is, the metabolic limits of the nervous system mentioned above [Clarke and Sokoloff, 1999, Lennie, 2003], and from experimental observations of performances during visual search tasks [Posner et al., 1980, Eriksen and Collins, 1969, Jonides, 1983]. While the area of the visual field which falls on the fovea of the retina is highly resolved and may explain the "overt" focus of attention, the manner in which "covert" attention is distributed over the para-foveal & peripheral regions, and the dynamics of its shifting to new areas have been subject to a variety of theories.

Of the first theories of visual spatial attention was the notion that attention was analogous to a 'spotlight' [Posner et al., 1980]. Such a model postulated that whatever area the attentional spotlight shone on was attended to (i.e. visually processed), and the rest of the visual field was left unattended (i.e. unprocessed). Evidence for selective attention was seen from experiments involving manual response times to LED flashes at expected, neutral, and unexpected locations on a display screen. While fixated on screen center, response times were significantly faster for flashes at expected locations (via a pre-cue) compared to response times to unexpected locations. The pre-cueing of attention

was termed 'orienting' and the process of selectively attending to objects in the visual field was termed 'detection'. The "covert" beam of attention moves independently of the retina, and as such the larger the distance to cover across the visual field the greater the time needed to adjust attention.

The spotlight model of attention was then expanded upon by adding to the analogy the quality of a 'zoom-lens' modulating the (quite literal) focus of attention [Eriksen and James, 1986]. The concept of a zoom-lens followed from experimental results noting the capacity of selective visual processing to spread over a large area of the visual field, or to be concentrated on smaller areas of the visual field [Eriksen and Collins, 1969]. Additionally, results of experiments where reaction time decreased when attention was oriented prior to visual search [Colegatef et al., 1973, Eriksen and Colegate, 1970] were expanded upon to note that visual reaction times were directly correlated to the reliability of a pre-cue to orient attention [Jonides, 1983]. As such, a two-process model of attention was hypothesized where attention shifted between two 'modes': one focused for when the pre-cue was reliable, and the other dispersed for when the pre-cue was unreliable (so all items needed to be searched in parallel).

These two modes were then theorized as being two extremes of a continuum of attentional focus 'levels' - much like a zoom-lens attachment on a camera [Eriksen and James, 1986]. Fundamentally, the model posits that processing capability and processing area are inversely related variables so that the scope of selective attention may vary in size, but attending to larger areas will reduce overall visual processing efficiency while attending to smaller areas will increase efficiency. Neuro-physiological evidence for the zoom-lens model has been seen using fMRI imaging of visual cortical areas and a pre-cueing stimulus to change the area of attention; an increase of area corresponded to an increase size of active cortical regions but a decreased level of neural activity [Müller et al., 2003]. The capability of the attentional system to flexibly allocate resources across the visual field in order to optimally configure processing resources raises a question as to how and why the system might favour certain configurations over others, and the manner in which it adapts to account for dynamic stimuli.

Behavioural tasks for probing visual spatial attention

A variety of tasks have been used to probe visual spatial attention, specifically 'covert' attention (as 'overt' attention can be investigated via pupil position tracking). Response times as a measure of covert attention have been frequently used due to the notion that pre-allocating visual processing resources to locations in the visual field facilitates the detection of a target in that location, hence a faster response time [Posner et al., 1980]. Tasks in which response times are used include visual search tasks where participants must search for a target object among distractor objects [Posner et al., 1980, Posner et al., 1984, Colegatef et al., 1973, Eriksen and James, 1986, Jonides, 1983]. Other methods for evaluating the distribution of visual spatial attention have measured attention by the ability to discriminate characters, rather than using target detection times [Khurana and Kowler, 1987, Lovejoy et al., 2009]. In these tasks, participants are presented with rows of alphanumeric characters and asked whether a specific character appeared in the array. Finally, besides visual search tasks, the response times to the abrupt onset of salient visual probes has been used to evaluate the allocation of spatial attention [Donkelaar and Drew, 2002, Khan et al., 2010, Seya and Mori, 2012]. As abrupt onsets of stimuli have been shown to capture attention in general [Yantis and Jonides, 1984], asymmetry in response times would be indicative of an existing bias in spatial attention prior to stimulus onset [Khan et al., 2010].

Visual spatial attention during motion tracking

During smooth pursuit tracking of motion, spatial attention has been shown to be directionally focused at locations ahead of target motion [Donkelaar and Drew, 2002, Khan et al., 2010, Seya and Mori, 2012, Chen et al., 2017]. The extent of the bias in attention ahead versus behind target motion as a function of target speed has shown conflicting results, with even low speeds of about 5 degrees per second producing a more pronounced bias located further ahead in some studies [Seya and Mori, 2012], but only showing this effect at higher speeds (greater than 10 degrees per second) in others [Donkelaar and Drew, 2002]. The reasons for the bias ahead of target motion has been likened to smooth pursuit (or target motion) acting as an explicit "endogenous"

(goal-directed) cue to orient spatial attention [Khan et al., 2010].

Visual motion prediction

Visual motion prediction refers to the ability of the nervous system to extrapolate the motion of an object and predict its future location. Transmission of signals through the nervous system naturally has delays, and so motion prediction is a necessary process in order to 'keep up' with objects in real-time [Nijhawan, 1994]. The role and capability of various structures within the visual system (specifically, the dorsal stream responsible for space localization) to extrapolate motion have been examined, and models of feedforward-feedback loops among structures such as the primary visual cortex and the middle temporal region to generate a consistent perception of motion have been proposed [Hu et al., 2022]. The manner in which these systems interact with visual attention systems remains to be explored further.

Motion extrapolation

Generally, the ability to predict the future position of an object effectively rests on the scheme which a system adopts to extrapolate the object motion. Furthermore, the timescale of extrapolation is also a crucial factor in the practical utility of predictive mechanisms - being able to predict the position of a baseball only once it crosses home-plate only results in strike-outs. As such, it is critical to understand what types of visual cues aid motion extrapolation, and for how long the information from those cues persist (i.e. how far in advance that information is useful) in driving visual motion extrapolation schemes. The concept that the motion history of an object distorts the perception of that object's instantaneous position can be seen via a "flash-lag" effect [Munger and Owens, 2004, Hubbard and Bharucha, 1988, Freyd and Finke, 1984, Nijhawan, 1994]. In these experiments, an object is shown to rotate or move in some predictable fashion for a period of time before disappearing and then reappearing. The participant is asked to judge whether the object reappeared in the same orientation it had when it disappeared. A reliable error has been observed that participants perceive

the object to have moved slightly beyond its disappearance point. The "flash-lag" effect has been attributed to a mechanism termed representational momentum (RM), where an object following a pattern of motion has an internal representation which continues to undergo forward dynamics despite the physical state of the object [Freyd and Finke, 1984].

The concept of a "pattern" is dependent on the participant's a priori expectations of motion, such as adherence to Newtonian physics [Hubbard and Bharucha, 1988, Nagai et al., 2002]. As such, objects moving in the direction of gravity relative to body position are perceived as having more forward displacement compared to those moving against the direction of gravity - a special case of RM specifically termed "representational gravity". The object's avatar or label may also sway expectations, such as having the motion stimulus resemble a rocket or a steeple, which are deemed to move (if at all) differently in the real world [Reed and Vinson, 1996]; participants demonstrated less representational momentum for steeple avatars compared to rocket avatars, especially when the implied direction of motion was upwards. Finally, beyond the effect of real-world exposures to movement, patterns such as periodicity may also be inferred from motion stimuli within the experimental paradigm [Verfaillie and d'Ydewalle, 1991]. Participants exposed to a modified flash-lag protocol with periodic rotational motion demonstrated no forward representational momentum at direction change-points, and when presented with periodic velocity fluctuations there were corresponding levels of forward extrapolation at velocity change-points [Verfaillie and d'Ydewalle, 1991]. In fact, in the absence of patterned (or predictable) motion, the forward shifts of perceived motion completely disappear [Kerzel, 2002]. In conclusion, representations of objects and their motion patterns which may guide motion extrapolation can be extracted from motion history or based in real-world contexts, and these patterns may manifest across movement orders.

Motion extrapolation in athletic populations

One of the first experiments investigating the "flash-lag" effect involved baseball athletes and examined the manner in which the baseball motion was extrapolated into the future to facilitate accurate catching [Nijhawan, 1994]. Ball sport athletes, especially

from sports involving the catching or interception of a ball (such as baseball), have since been involved in various motion extrapolation paradigms, including the go/no-go task [Sokhn et al., 2019] and time-to-contact tasks [Spering et al., 2011, Fooken and Spering, 2019]. Notably, the superior predictive ability in athletes may also arise due to contextual knowledge learned from practice which would lead to more effective visual strategies for detecting information which would facilitate prediction. For example, knowing how to "read" the body mechanics of a pitcher (i.e. what body parts to look at and when) constrains the possibilities of ball trajectories, thus facilitating the prediction of future ball position even before the ball is thrown [Abernethy and Russell, 1984]. However, even following the release of the ball (and hence a lack of contextual body cues), experienced athletes are better able to predict trajectory from the first 150 milliseconds of ball flight, and thus initiate a saccade towards the predicted future location quicker than less experienced athletes [Land and McLeod, 2000]. Furthermore, not only does an athlete's level of play reflect performance in motion extrapolation tasks, but also the regular practice of sport can improve motion extrapolation performance over time [Kida et al., 2005].

Predictive coding of visual motion

The predictive 'coding' of sensory stimuli revolves around the theory that the brain is actively employing an internal model of the surrounding environment [Francis and Wonham, 1976, Millidge et al., 2021]. The internal model is used to predict the future state of the environment and the body to overcome delays in neural processing, and is constantly updated by acquiring new sensory input [Rao and Ballard, 1999]. In a predictive coding framework, the observer's primary objective is to minimize the error between the internal model predictions and the sensory information being observed, and in the process of this minimization the observing system effectively adapts its internal model. By considering the brain as a predictive coding machine, a computational model revolving around the hierarchical organization of cortical layers is proposed [Friston, 2003], where prediction occurs simultaneously across several levels which interact with each other via several feed-forward and feed-back loops [Clark, 2013] . From the compu-

tational framework, the function of the feed-back paths allow for predictions made at higher levels of the hierarchy to be compared to inputs at lower levels, while feed-forward pathways allow for the errors in prediction to be fed back into the system to improve model calibration [Hogendoorn and Burkitt, 2019]. Due to processing delays, information reaching the "higher" (or deeper) levels of the hierarchy must compensate by making longer range predictions and therefore are likely to encode longer range (or global) features, while "lower" levels have less time to compensate for and likely encode local characteristics [Hogendoorn and Burkitt, 2019, Millidge et al., 2021]. Studies to date have both developed mathematical frameworks and discovered biological evidence to support the notion of multi-layered hierarchical neural circuitry driving the prediction of visual motion [Hogendoorn and Burkitt, 2019, van Heusden et al., 2019, Berry et al., 1999, Jehee et al., 2006, Spratling, 2012].

Neural mechanisms of predictive coding

The functional architecture of the visual system has often been observed as resembling a hierarchy of connections [Hubel and Wiesel, 1962, Fukushima, 1988, Riesenhuber and Poggio, 1999]. In the experiments of Hubel & Wiesel [Hubel and Wiesel, 1962], receptive fields of neurons in the lateral geniculate nucleus (LGN) were relatively small and responded optimally to point light stimuli, whereas receptive fields of neurons in layer IV of the primary visual cortex (V1) responded optimally to lines of a particular orientation. As the LGN is a primary source of afferents to V1, the implication is that V1 cells integrate these afferents to build relatively higher-order features of the visual input. As visually processed features are passed iteratively onwards to other cortical structures, the afferents signals encoding those features are combined further to produce even more detailed or complex features such as contours for object recognition in V4 [Riesenhuber and Poggio, 1999] or optical flow for motion processing in area MT [Solari et al., 2015].

The capability of motion extrapolation to occur beginning at the level of the retina has been demonstrated, where retinal ganglion cells ahead of a moving target exhibit an increase in firing rates irrespective of motion direction and across a range of speeds, al-

though extrapolation deteriorates at sufficiently high speeds or insufficiently high contrast values [Berry et al., 1999]. Cells in the LGN also have shown similar "pre-activation" corresponding to the encoding of future positions of a moving contour which can be specifically linked to feedback projections from V1 [Sillito et al., 1994]. The receptive fields of V1 cells in cats have also shown "pre-activation" sensitive to motion speed and direction, and a peak activation that was quicker for dynamic rather than stationary stimuli [Jancke et al., 2004]. Indeed, for motion prediction mechanisms to be of any use there must be advantages in processing speed so that predictions induce responses quicker than if relying on sensory stimuli alone - evidence for these processing advantages have been observed via electroencephalographic recordings during a motion anticipation task [Blom et al., 2020].

In addition to predictive models generated solely by forward dynamics from internal models and backwards error minimization schemes, the use of motor plan to help predict into the future has also been explored via the concept of "efference copies". An efference copy is a duplicate of the motor plan (the oculomotor plan, in the case of eye movements), which can be used by internal representations of space to maintain a stable perception of the world [Stark and Bridgeman, 1983]. During visual motion prediction tasks, where participants must predict the time-to-contact or final location of a target after partial occlusion of its trajectory, smooth pursuit of the target resulted in better performance compared to fixation strategies [Spering et al., 2011, Fooken and Spering, 2019]. The performance improvement associated with smooth pursuit over fixation remained even with equivalent retinal stimulation, and more accurate eye movements were indicative of better motion prediction ability [Spering et al., 2011]. These observed advantages point towards the use of an efference copy of the oculomotor commands during smooth pursuit as additional latent variables aiding predictive coding models alongside sensory input from the visual processing of motion.

Mathematical framework

In Bayesian statistics, the concept of inference refers to the testing and updating of the probability that a hypothesis is true, given prior probability distributions on that hypothesis and sampled data to generate a likelihood function [Berrar, 2018]. The process of inference is summarized using Bayes' theorem:

$$\mathbf{P}(H|E) = \frac{\mathbf{P}(E|H) \cdot \mathbf{P}(H)}{\mathbf{P}(E)} \tag{1}$$

Where $\mathbf{P}(H|E)$ is the posterior probability of the hypothesis being true given the sampled data, $\mathbf{P}(H)$ is the prior hypothesis probability distribution, $\mathbf{P}(E|H)$ is the likelihood of the sampled data occurring given the prior hypothesis, and $\mathbf{P}(E)$ is the likelihood of the sampled data independent of the hypothesis. Naturally, the computation of $\mathbf{P}(E)$ is typically intractable as it requires the computation of the probability that the sampled data occurs in any latent (i.e. unobservable) state; to solve the problem, variational Bayesian inference is employed whereby the posterior probability is computed via minimizing the relative entropy between it and variations on an arbitrary probability distribution [Beal, 2003].

The concept of an error reduction system for predictive coding can be reformulated as variational Bayesian inference [Friston, 2003]. Predictability and information are inversely linked [Shannon, 1948], as outcomes from a purely predictable process produce no new information about that process whereas outcomes from a purely random process are always producing new information. A model needs to be assumed in order to "predict" anything, which corresponds to the prior hypothesis probability distribution, P(H), in Bayes' theorem. New information sampled by the system can then be used to generate a likelihood function of that data, given the prior distribution. To solve for the posterior probability distribution (i.e. "update" the model to better reflect the sampled data) an optimization scheme is run so that the amount of variational "free-energy" (relative entropy) is minimized [Friston, 2003].

Given a mechanism for inference and learning, a second key aspect of predictive coding

is a hierarchical structure which integrates the individual learning of multiple components across the system to produce a coherent perception of motion [Hogendoorn and Burkitt, 2019]. The function of components in higher levels of the hierarchical circuit is to predict the level of activation in components in lower levels of the hierarchy. The predicted activations can then be held by the corresponding lower components, compared to sampled values at the appropriate time, and error in predicted and sampled data can be fed back upstream to the higher level components [Rao and Ballard, 1999]. However, a key drawback of the classical conception of the hierarchical model is that since processing delays naturally occur in the transmission of information across levels, the time delay involved in moving predictions across feed-forward and feed-back pathways necessarily introduces increased prediction error as well as incoherence due to each level of the hierarchy representing the object at a different point in time [Rao and Ballard, 1999]. Hence, one extension to the classical model involves the modulation of prediction signals along feedback connections to be modulated to account for transmission time, hence feeding back a predicted activation of the lower component at the instant of arrival rather than the instant of relay. Next, in order to temporally bind components across the hierarchy, a forward extrapolation scheme must also be used whereby the prediction errors in the forward loops are modulated to account for the time of transmission to higher levels, so that higher components are producing predictions on the activation levels of lower components given errors modulated to reflect instantaneous values [Hogendoorn and Burkitt, 2019].

Predictive coding frameworks can be simplified into linear, state-space models such as the Kalman filter [Baltieri and Isomura, 2021]. Such recursive estimators implement two primary steps: a prediction step iterating the current state estimation using an internal dynamic model, and a correction step which adjusts the predicted state estimate based on incoming measurements. Furthermore, Kalman filters can themselves be arranged in hierarchical networks to combine multiple state estimations into a single optimized state estimate. For example, in modelling mechanisms underlying smooth pursuit, two Kalman filters were employed to independently represent the nervous system's measurement (i.e. from retinal slip) and prediction (i.e. from an internal memory model) of the pursuit tar-

get's position [de Xivry et al., 2013]. The two estimations were then optimally combined using a Bayesian integrator to generate a final estimate of the target's state which could be used to drive eye movements. Beyond the explicit prediction of motion or position, Kalman filters have also been used to model the nervous system's estimation of its own error in tracking future target positions [Nachmani et al., 2020]. Specifically, the state covariance matrix estimated by the Kalman filter gives an indication of the confidence of the state prediction, which can be used to generate a probability of the target falling outside the fovea (i.e. a predicted future error) and thus trigger a catch-up saccade during smooth pursuit, if appropriate. In conclusion, state space models and recursive estimators have been shown to be effective in characterizing and modeling the manner in which the nervous system filters noisy measurements, predicts future motion, and controls eye movements.

Chapter II: Research questions & hypotheses

Research question 1

During visual motion tracking, is the distribution of visual spatial attention about the pursuit target dependent on motion predictability?

Hypothesis 1

I hypothesize that when visually tracking a target with predictable, uniform motion there will be a bias in visual spatial attention ahead of the pursuit target motion. When tracking a target with unpredictable motion, attention will be evenly distributed about the target motion (no difference between ahead and behind). Finally, when tracking a target with intermediate motion predictability will result in attention being distributed with some bias ahead of target motion, but not as much as in the case of uniform motion.

Research question 2

Is there symmetry in bias accumulation relative to changes in motion predictability? That is, does the time needed for bias to accumulate when viewing uniform motion following random motion equate to the time needed to de-accumulate bias when viewing random motion following uniform motion?

Hypothesis 2

I hypothesize that the time for the accumulation of bias when tracking a target moving from low-to-high predictability is higher than the time for the de-accumulation of bias when tracking a target moving from high-to-low predictability. As such, manual response times to probes located ahead versus behind target motion will take longer periods of uniform motion to be significantly different for bias accumulation compared to de-accumulation.

Research question 3

Can state space models be used to infer the dynamics of visual spatial attention from eye movement dynamics?

Hypothesis 3

While the research question is exploratory, I expect that state estimators can model visual spatial dynamics with some effectiveness as Kalman filtering has previously shown to model smooth pursuit and catch-up saccade dynamics. Assuming that target motion predictability (and hence, eye movement complexity) are linked to visual spatial attention (based on research question #1), the extension of such state space models to capture non-linear features, such as asymmetrical accumulation (based on the results of research question #2) should be feasible. In particular, the model should aim to take a target motion time series as input and produce the corresponding time series of visual response time bias as output. Using normalized values of reaction time bias, it is expected that a mean absolute error (MAE) of less than 50% is achievable (i.e. model prediction accuracy of attentional bias occurring is better than random chance).

I hypothesize that eye-movement acceleration can drive the modelling of attentional bias based on the idea that visual attention is biased ahead of target motion during linear pursuit due to the anticipation of future target location [Khan et al., 2010]. When target motion is un-predictable, it is hypothesized that visual tracking error will be high and hence there will be a need for frequent, high-acceleration eye-movement corrections (typically, catch-up saccades) to maintain tracking accuracy. Conversely, predictable target motion would mean a low occurrence of high-acceleration eye movement events. As such, the frequency and magnitude of eye acceleration events may be able to inform of the predictability of target motion, and hence the amount of bias in visual attention.

Chapter III: Methodology

Overview

Participants performed a dual-task paradigm involving smooth pursuit motion tracking and a visual spatial attention task. Smooth pursuit target motion was constrained to the horizontal axis. Visual attention was measured as a manual response time to a probe dot appearing on screen at random times and locations. To address research question #1, a target was programmed to move at a fixed level of motion predictability for up to 2400 ms, and attention was probed to determine the effects of motion predictability on response time bias. To address research question #2, some trials were presented where target motion predictability changed after 2400ms, and attention was probed to determine the accumulation or de-accumulation of bias in visual spatial attention. To address research question #3, a state-space model, taking motion inputs and producing normalized response time bias outputs, was evaluated with a 5-fold cross-validation using 400 motion trials across all participants.

Data collection

The stimulus was presented on a ViewSonic V3D245, with a resolution of 1920x1080 pixels, $120Hz$ refresh rate, and physical dimensions of $52cm$x$29.2cm$ (37 $^{pixels}/_{cm}$). Eye movements were collected using the Eyelink 2 (SR Research Ltd., Ottawa, ON, Canada) and recorded binocular gaze location from pupil-corneal reflection at $250Hz$. Participants were in a chin-rest to stabilize the head during collection, and infrared trackers connected to the Eyelink system were used to provide a head-tracking signal to correct for any further small movements of the head. The chin-rest was placed at a distance of $53cm$ from the monitor so that 1° of visual angle is equivalent to ∼40 pixels on screen; seating and chin-rest positions was adjusted so that the participant's gaze was centered both horizontally and vertically on-screen. Manual response times were registered via a standard PC mouse (polling rate of $125Hz$).

Sample population

A total of 20 participants between the ages of 18-35 were recruited, grouped as high-performance athletes (n=10; 5 male, 5 female) and healthy, young adults (n=10; 5 male, 5 female). High-performance athletes were considered to be individuals competing in a dynamic team-sport at the varsity level of play or higher. Athletes have been shown to exhibit higher and more consistent levels of dynamic visual acuity [Yee, 2017], potentially providing a population with similar experience levels and practiced visual strategies.

All participants performed a series of visual assessments prior to data collection including: static visual acuity, dynamic visual acuity, and stereo-acuity. Additionally, age and sex were recorded; all personal identifiers were stored separately for confidentiality purposes. All participants completed an informed consent form prior to participation in data collection. The research protocol was approved by the University of Waterloo Research Ethics Board under study #43502.

Sample size

In previous research studies investigating spatial allocation of attention using response times to probes, population sample sizes ranged from 4 to 9 participants [Donkelaar, 1999, Lovejoy et al., 2009, Khan et al., 2010]. Pilot data was collected from 6 participants to determine the effect size (Cohen's d) of the bias between response times ahead versus behind target motion during smooth pursuit; the resulting effect size was greater than 1.5 (d=2.11). A power analysis was then run using G-Power [Faul et al., 2007] with the following input parameters for a paired sample t-test: one-tail, effect size of 1.5, alpha level of 0.05, power of 0.95. The sample size estimate was 7 participants. A sample size of 20 individuals in total (10 in each group) was used to be able to compare groups and account for possible participant data exclusion during analysis due to poor data quality.

Exclusion criteria

Individuals with a history of concussion were excluded due to possible lingering effects on cognitive ability [Hurtubise et al., 2016]. Individuals with clinically diagnosed atten-

tion deficiency disorder were also excluded. Additionally, participants were screened for handedness using the Edinburgh inventory [Oldfield, 1971], and only right-handed individuals were recruited to control for possible sources of variances arising from differences in lateralization within the brain which possibly affect performance on visuo-spatial tasks [Gordon and Kravetz, 1991].

Methods

Motion complexity & persistence

To quantify the predictability of motion, a metric of complexity derived from information theory was used to define a discrete random walk model which generated the target movement sequences. A random walk is a stochastic process meant to model the movement of a free particle in space - that is, Brownian motion. A discrete random walk considers the states of a particle to be discrete, rather than continuous, and can approximate the continuous random walk by taking the time duration of each state as a limit approaching zero [Kac, 1947]. However, randomness can also be considered as a state of maximal entropy [Shannon, 1948], thereby allowing for a parametric approach to defining a varying level of randomness.

Entropy is defined as the level of uncertainty in the state of a system and can be computed as follows:

$$H = -\sum_{i=0}^{n} p_i log(p_i) \tag{2}$$

Here the system is modelled by n states, and p_i is the probability of the system being in the i-th state. If there is equal probability to be in any state (i.e. the system is governed by a uniform distribution), then entropy is maximal; of course, equal probability of any state is a synonymous description of pure 'randomness'. By contrast, if the probability of a state is 1, the probability of all other states is 0; the logarithmic function itself evaluates to 0 at 1, and so the entropy is trivially zero as well. In this case, the system is purely deterministic - there is no uncertainty in the state at all.

The notion of random walks and entropy can be combined via Markov-chain models.

A Markov chain considers the probability of the next state of a system as determined by only the previous N states of that system [Gagniuc, 2017]; the value of N denotes the order of the Markov chain (or system 'memory'). By modelling a random walk as a first order Markov-chain, one can imagine a walker moving to the left on a line who flips a coin at each step to decide whether to continue in the current direction of motion (heads) or change direction (tails). Clearly, with a fair coin toss the probability of maintaining or changing direction is a 50-50 chance. The resulting motion of the walker would thereby resemble purely random motion along the line. However, consider an unfair coin toss, where both sides of the coin are heads - then the walker can only ever continue in the same motion direction.

A Markov-chain model of random walks is not simply constrained to the purely random and purely deterministic conditions; intermediate levels of motion randomness can be achieved by setting the state transition probability to values in between 0 and 1 (and therefore having recurrence probability values between 1 and 0, respectively). The 'randomness' in a motion can then be described in a more generic manner as its 'complexity', with the following implications:

- motion complexity can be quantified by the entropy of the stationary distribution governing transition and recurrence probabilities. Note that a symmetric definition of the random walk model results in a single parameter (state transition probability, P_T) governing transition and recurrence in all states. Therefore, the transition matrix of the Markov chain reduces to a stationary distribution, and complexity is equivalent to Shannon's entropy as defined by Equation 2.

- maximally complex motion ($P_T = 0.5$) is analogous to purely random ("unpredictable") motion, while minimally complex motion ($P_T = 0$) is analogous to purely deterministic ("predictable") motion.

- other levels of motion complexity, where entropy is between 0 and 1, are indicative of the level of 'persistence' in a motion; the lower the entropy value, the more persistence exists.

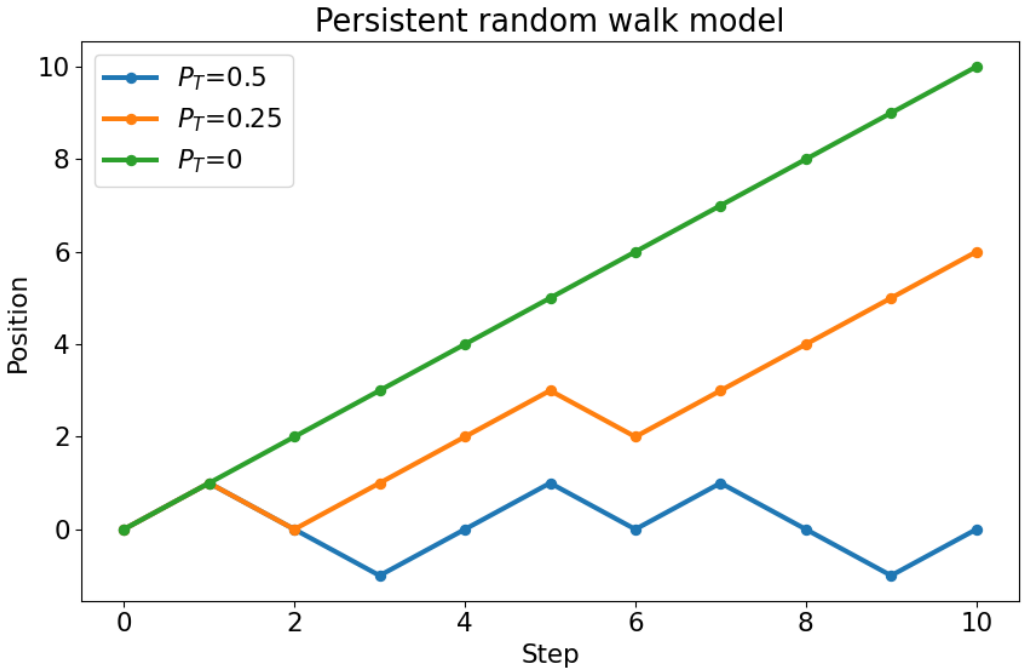

Figure 1: **Persistent random walk model:** Simulation of a persistent random walk using a Markov chain model at varying transition probability (P_T) levels. For P_T levels close to 0, the walk results in straight line, uniform motion, whereas when P_T levels increase to 0.5 the uniform motion begins to fluctuate increasingly until it resembles pure Brownian motion at P_T=0.5. Increasing P_T past 0.5 and into the range of [0.5, 1] produces a predictable oscillatory motion pattern (predictable state switching)

Stimulus generation

The target appeared as a white dot (0.5 degree diameter), and moved at a constant velocity of 10 $^{\circ}/_{sec}$. Target size and velocity was selected to be comparable to similar studies [Khan et al., 2010], as well as from consideration of pilot data demonstrating a bias in manual response times from these target parameters. Target motion was constrained to the horizontal axis to avoid confounding different systems of control for vertical and horizontal pursuit [Rottach et al., 1996].

A persistent, Markov-chain random walk model was used to generate target motions of varying complexity levels by changing transition probability (P_T) values. System states were defined as target velocity, either moving at a constant positive rate (+10 $^{\circ}/_{sec}$) or a constant negative rate (-10 $^{\circ}/_{sec}$).

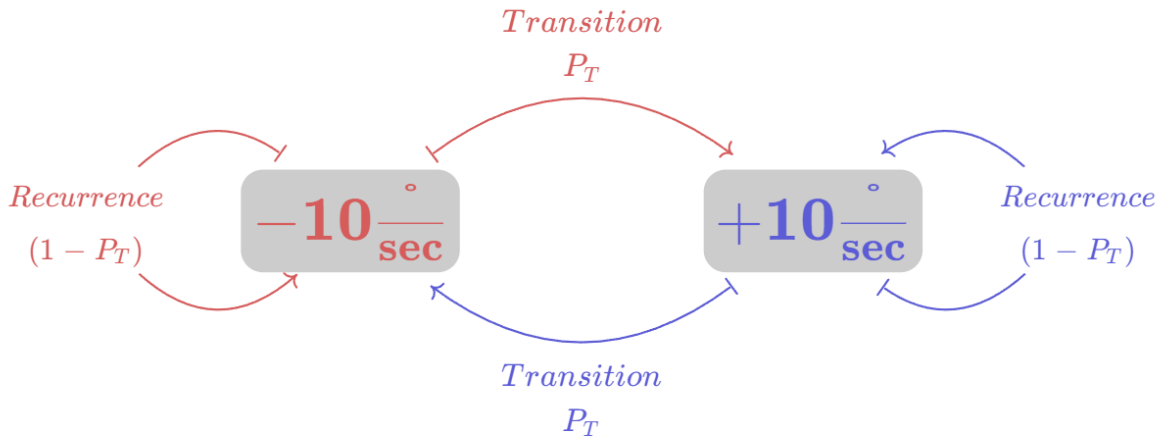

Figure 2: **Random walk Markov chain:** First-order Markov model of the persistent random walk. A single parameter, P_T, governs the transition probability of the system. System states are symmetric so that the transition probability resembles a "stability" index (i.e. persistence) of the entire system regardless of its current state (in this case, its angular velocity).

In order to address research question #1, target motion was generated using the persistent random walk model by fixing the transition probability at values of $P_T = 0, 0.25, 0.5$. The duration of a single velocity state (or 'step' of the motion) was 300 ms long generated using 8 iterations of the model.

In order to address research question #2, target motion where $P_T = 0, 0.5$ had a second phase of motion consisting of 8 iterations each lasting 300 ms where $P_T = 0.5, 0$, respectively. As such, these trials consisted of either predictable-to-unpredictable motion ($P_T = 0 \rightarrow 0.5$) or unpredictable-to-predictable motion ($P_T = 0.5 \rightarrow 0$) lasting 4800 ms in total (16 iterations of 300 ms each).

In summary, three blocks of target motion were presented to the participant in a random order:

1. 2400 ms of predictable motion → 2400 ms of unpredictable motion
2. 2400 ms of semi-predictable motion
3. 2400 ms of unpredictable motion → 2400 ms of predictable motion

Probes flashed ahead or behind target motion in the first 2400 ms of motion were used

to assess visual attentional bias due to motion predictability for research question #1, while probes flashed in the latter 2400 ms of motion types 1 & 3 were used to assess the adaptation of attentional bias for research question #2.

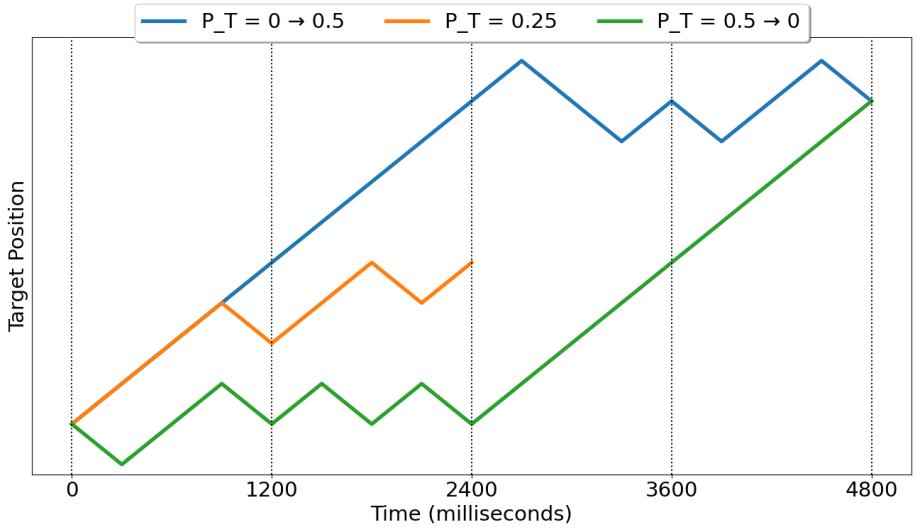

Figure 3: **Target motion generation:** Predictable-to-unpredictable (blue), semi-predictable (orange), and unpredictable-to-predictable (green) stimulus motion types. The vertical dashed line distinguishes the first phase of motion (2400 ms) from the second phase of motion. Vertical dashed lines separate the 4 probe time windows (1-1200 ms, 1201-2400 ms, 2401-3600 ms, 3601-4800 ms).

Probe generation

Bright red visual probes (0.25 degree diameter) were flashed (10 ms) randomly, either 2 or 6° ahead or behind target location. A probe eccentricity of 2° has been shown to be most sensitive to attentional bias during smooth pursuit both in previous literature [Khan et al., 2010, Donkelaar and Drew, 2002] and from pilot data sessions. However, the 6° eccentricity were also used to probe any effects of motion predictability or motion dynamics on attention in the peripheral visual field. For each trial, the probe was flashed in one of the following time windows: 1-1200 ms, 1201-2400 ms, 2401-3600ms, or 3601-4800ms.

Response time collection

Previously, both saccadic and manual response times have been collected to measure visual attention. However, saccadic response times are affected by smooth pursuit based on saccade target location relative to smooth pursuit target location [Khan et al., 2010]. The interference effect of smooth pursuit on saccade response time is theorized to stem from a common control process governing both types of eye movements [Xivry and Lefèvre, 2007]. Saccades to probe locations relatively nearly ahead of target motion are suppressed as an optimization strategy, as smooth pursuit is deemed to reach the probed location in a similar amount of time as a saccade [Brouwer et al., 2002]. Therefore, to avoid possible confounds on probe response times, manual response times via a mouse button press were recorded.

Experimental protocol

The experimental protocol followed the overall procedure below:

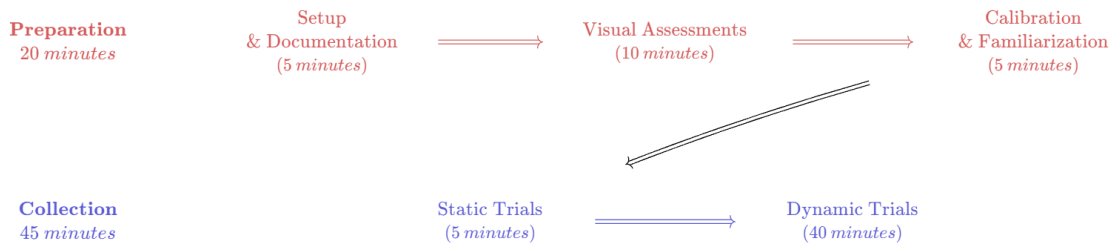

Figure 4: **Experimental protocol overview:** Schematic outline the flow of procedures during a data collection session

Visual assessments

Following protocol explanation and informed consent, participants began data collection by completing a set of visual assessments. Static visual acuity was tested using an Early Treatment Diabetic Retinopathy Study (ETDRS) letter chart [Bailey and Lovie-Kitchin, 2013]. Dynamic visual acuity was tested using a monitor displaying an 'E' moving either randomly or horizontally. The 'E' presented itself on a television monitor in various orientations (up/right/left/down) and moving at a speed of $\sim 15^{\circ}/_{sec}$ and the participant indicates

orientation via a button press [Hirano et al., 2017]. For both static and dynamic visual acuity assessments, participants were seated 4 meters from the assessment plane (letter chart or monitor, respectively) and acuity scores were determined per line.

Calibration & familiarization

A 5-point calibration using pupil-corneal reflection on both eyes was employed. Validation levels were recorded and calibration repeated if the error in either eye was greater than 0.5 degrees. Following calibration, a maximum of 32 familiarization trials occurred where participants were acquainted with the dual-task paradigm. Motion generated for the familiarization trials employed a random walk model as described in stimulus generation, with transition probability levels randomly set to $P_T = 0, 0.25, 0.5$ and probes randomly located 2°, 6° ahead or behind target motion. None of the pre-programmed motion for the familiarization trials were replicated for the experimental trials.

Static visual spatial attention

In order to map the distribution of visual spatial attention in the absence of motion tracking, 40 trials were presented where the participant fixated a target on the screen center and manually responded to probes 2° or 6° to either the right or left of screen center. The resulting static visual spatial attention distribution was then used for comparison to the dynamic visual attention distribution and to normalize response time data for each subject.

Dynamic visual spatial attention

For each participant, a total of 320 trials of the dynamic visual spatial attention task were collected, composed of 128 trials of predictable-to-unpredictable motion ($P_T = 0 \rightarrow 1$), 64 trials of semi-predictable motion ($P_T = 0.25$), and 128 trials of unpredictable-to-predictable motion ($P_T = 1 \rightarrow 0$). Trials were split into blocks of 32 trials, for a total of 10 blocks and all block orders were randomized. Within each block, the probe location (ahead or behind target motion), the probe distance (2° or 6°) and the probe time was

randomized on a trial-by-trial basis.

Data processing

Eye movement data, stimulus motion, probe events and manual response events were exported via DataViewer software (SR Research Ltd., version 4.3.1) for subsequent offline processing and analysis. Custom Python scripts were used to extract saccades, filter eye movement traces, and determine smooth pursuit accuracy as follows:

1. Saccade detection used the Engbert-Kliegl algorithm, with a position threshold of 0.1°, a velocity threshold of $30^{\circ}/_{sec}$ and an acceleration threshold of $800^{\circ}/_{sec^2}$ (Engbert & Kliegl, 2003; Engbert, 2006).

2. Filtering eye position time-series data involved replacing detected saccades with linear interpolations prior to applying a low-pass, second-order Butterworth filter set to a cutoff frequency of $40Hz$.

3. To ensure consistent smooth pursuit accuracy, trials where the mean eye-target distance exceeds 2° (from 200 ms following target motion onset to probe appearance) were discarded.

Static trials (used to gauge baseline response time) where manual response times were less than 100 ms were discarded due to physiological in-feasibility [Baars and Gage, 2018] and based on previous studies demonstrating simple visual response times in competitive athletes around 130 ms [TP et al.,]. For dynamic trials, manual response times less than 200 ms were discarded, corresponding to choice reaction times in competitive athletes of around 260 ms [TP et al.,]. Additionally, outlier response times will be dropped from the analysis, with outlier bounds determined for each subject as: $[1^{st}\,quartile - 1.5 * IQR$, $3^{rd}\,quartile + 1.5 * IQR]$, where IQR is the inter-quartile range of all response times. Finally, trials where blinks occurred within 100 ms of probe appearance were discarded.

The above data exclusion criteria stem from similar practices employed in previous research evaluating smooth pursuit quality for attention and motion prediction [Lovejoy et al., 2009, Khan et al., 2010, Spering et al., 2011].

Data analysis

Data normalization

Manual response times were normalized linearly by removing the mean baseline response time, as collected during the static protocol.

Visual spatial attention bias

To obtain a measure of visual spatial attention bias $\mathbb{B}$, the difference between normalized response times behind (RT_B) versus ahead (RT_A) target motion was computed, for each probe time window.

$$\mathbb{B} = RT_A - RT_B \tag{3}$$

Visual tracking error

To obtain a measure of visual tracking error, the Euclidean distance between the gaze and target position was used.

$$tracking\,error = \sqrt{(target_x - gaze_x)^2 + (target_y - gaze_y)^2} \tag{4}$$

Statistical analysis

Statistical analyses were primarily be conducted in R [RCoreTeam, 2021]. For all tests, significance levels were assigned as $p < 0.05$.

Normality assumptions characteristic of parametric statistics were tested by a Shapiro-Wilk test. A log transformation of data was used to correct for any deviations from normality.

Descriptive statistics were calculated for both normalized response times and visual tracking errors for each combination of a) participant group, b) probe location groups (ahead, behind), c) probe distance (2°, 6°), d) probe time window (1-1200 ms, 1201-2400 ms, 2401-3600 ms, 3601-4800 ms), and e) motion complexity level ($P_T = 0, 0.25, 0.5$).

To address research question #1, 2 three-way mixed analysis of variances (ANOVAs)

were conducted to determine the effect of participant group (between), motion predictability (within), and probe location (within) on normalized response times, one for each probe distance (2°, 6°). Additionally, to establish how motion predictability may also affect visual tracking accuracy, a two-way mixed ANOVA will be performed to determine the effect of group (between) and motion predictability (within) on tracking error.

To address research question #2, 4 two-way ANOVAs will be performed to determine the effect of group (between) and probe time window (within) on response time bias ($\mathbb{B}$), one for each of the following conditions:

- ANOVA #1: probe distance of 2°, predictable-to-unpredictable motion
- ANOVA #2: probe distance of 2°, unpredictable-to-predictable motion
- ANOVA #3: probe distance of 6°, predictable-to-unpredictable motion
- ANOVA #4: probe distance of 6°, unpredictable-to-predictable motion

Modeling visual spatial attention during motion tracking

To address research question #3, a state space model was trained and tested on the experimental visual tracking and response time data. The inputs of the model were the stimulus target motion files and the outputs were the corresponding predicted time series of response time bias (on a normalized scale of 0 - no bias, to 1 - maximal bias).

The state space model relied on two main components: a) a Kalman filter to simulate smooth pursuit of the target (blue loop in Figure 5), and b) a non-linear adaptation of the "leaky integrator" (causal filter) to transform simulated pursuit dynamics into normalized response time bias (red loop in Figure 5).

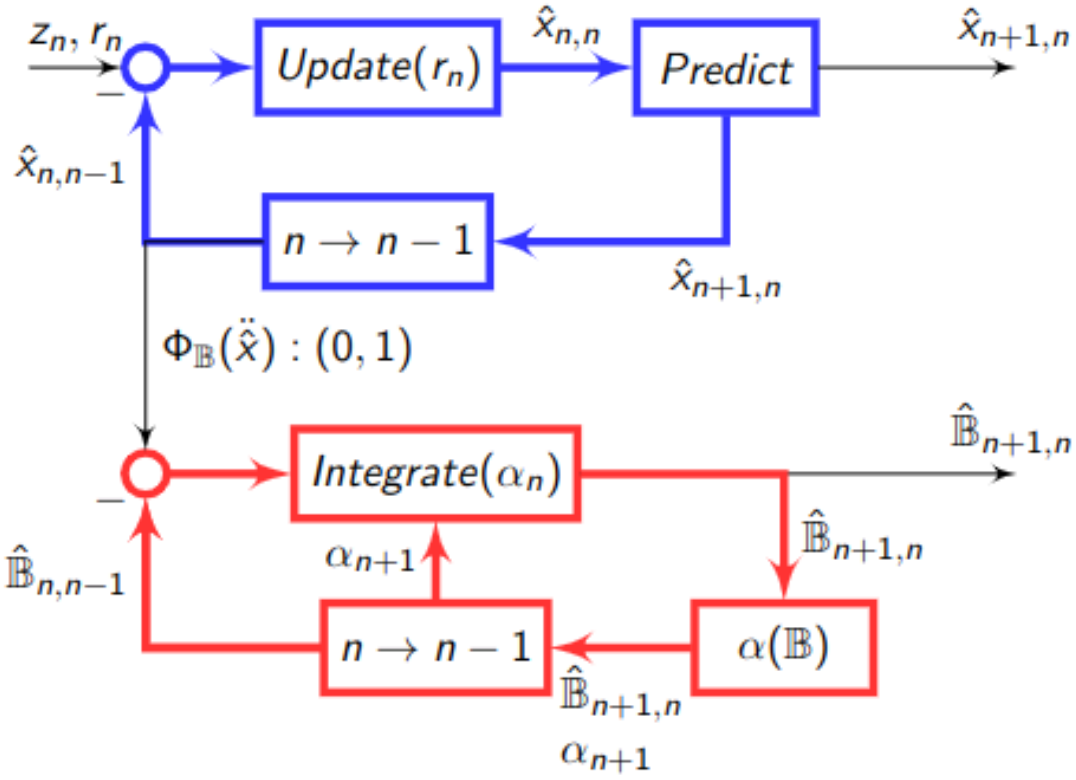

Figure 5: **State space model overview:** Schematic outline of model: components in blue represent a Kalman filtering scheme for motion tracking, while components in red represent a leaky integrator for the accumulation of attentional bias. Model inputs are measured target position and precision (z_n, r_n); the Kalman filter outputs predicted future target position $(\hat{x}_{n+1,n})$ for oculomotor actuation; the predicted target acceleration $(\ddot{\hat{x}})$ is scaled and passed onto the leaky integrator to output a normalized attentional bias value $(\hat{\mathbb{B}}_{n+1,n})$

Kalman filter

A single Kalman filter was used to simulate smooth pursuit performance in response to target motion. To simulate the lag in eye-movements to target motion, the Kalman filter was tuned to the visual tracking performances by adjusting the measurement uncertainty covariance matrix (R); larger uncertainties imply longer delays in eye movement response (slower accumulation of evidence), while smaller uncertainties imply a quicker eye movement response to target motion (Figure 6).

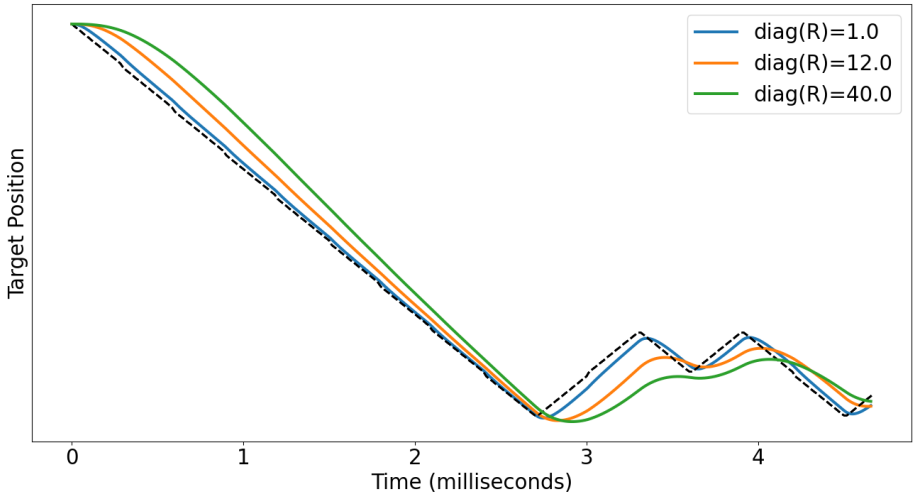

Figure 6: **Kalman filter simulation:** Effect of measurement error covariance matrix (R) on Kalman filter performance. Target motion (filter inputs) are represented as the dashed black line. Larger values on the diagonal of R ($diag(R)$) slow the eye movement response to target motion.

The covariance matrix was determined by adjusting the diagonal scalars of (R) and comparing the mean absolute error between the filter outputs and eye position data. The matrix R producing the lowest error across all trials was used as a hyper-parameter in the model. The output eye movement acceleration from the Kalman filter was then re-scaled into a range of [0, 1] using a logistic function with mean set at the third quartile of all simulated eye-movement data, so as to distinguish periods of recurring high-acceleration moments.

Non-linear leaky integrator

A "leaky integrator" is a linear filter used to estimate the state of a system based on a single parameter (α) which scales the residuals between incoming measurements and previous state estimates. Small values of α cause slow "leakage" of the system's memory of its previous state and hence delayed adaptation to incoming measurements, while large α values cause quicker memory "leakage" and therefore fast adaptation:

$$\hat{r}_k \leftarrow x_k - \hat{x}_{k-1}$$
$$\hat{x}_k \leftarrow \hat{x}_{k-1} + \alpha \cdot \hat{r}_k \tag{5}$$

Here r_k is first computed by taking the difference between the incoming state measurement (x_k) and the previous state estimation ($\hat{x}_{k-1}$. Then the new state estimation is achieved by adjusting the previous estimation according to the amount of confidence in state measurement versus state observation (i.e. the amount of leakage, α).

While the α value is typically set at a constant value, a non-linear adaptive implementation of the filter can be achieved by setting α as a function dependant on the current system state estimate. Such a scheme would allow for the integrator to have different rates of accumulation and de-accumulation, which may be necessary based on the results of research question #2. A suitable α function is the sigmoid function, as it can operate as a constant function if the rate parameter is set to 0 while increasing the sigmoid rate parameter approaches a step function for distinguishing accumulation from de-accumulation (Figure 7).

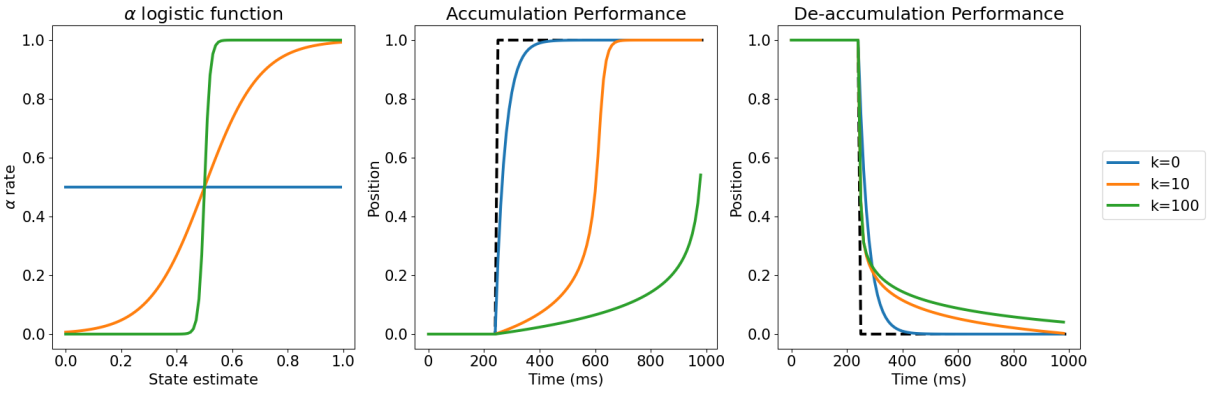

Figure 7: **Assymetric "leaky integrator":** Effect of rate parameter k of the α sigmoid function (**left**) and the corresponding integrator performance for accumulation (**middle**) and de-accumulation (**right**). Setting k=0 (blue) results in the constant function, hence the symmetric rates of accumulation and deaccumulation; increasing the value of k results in a slower accumulation rates compared to de-accumulation rates.

The optimal parameters for the α sigmoid function were learned by evaluating model

outputs to normalized experimental response time bias values using the range of sigmoid rates: [0, 128]. Model performance was quantified as mean absolute error between model outputs and experimental data.

Model dataset

In order to evaluate model performance, trials of experimental data were selected where trial duration lasted at least 4200 ms, in order to capture eye tracking and response time data across dual-phase motion (see). From each of the 20 participants, 20 such trials were randomly chosen, such that 10 trials had predictable-to-unpredictable motion and 10 trials had unpredictable-to-predictable motion, for a total of 400 trials. From these 400 trials, eye movement data was extracted in order to tune the measurement covariance matrix of the Kalman filter and determine acceleration re-scaling thresholds (); the resulting values were used as hyper-parameters for the rest of the model. Response time data was aggregated, normalized, and then interpolated using a nearest-value scheme in order to produce a response time bias time-series for evaluating instantaneous model outputs.

Model evaluation

The initial dataset (400 trials) was split into a training set (75%, 300 trials) and a testing set (25%, 100 trials). The training set was split into 5 folds of 60 trials each, where 4 folds (240 trials) were used to train the model parameters and 1 fold (60 trials) was used to validate the model parameters. All data sub-sets ensured an even split between predictable-to-unpredictable motion trials and unpredictable-to-predictable motion trials. Five iterations of training & validation occurred so that each fold was used as a validation set once, thus producing 5 sets of trained model parameters. The optimal set of parameters was then chosen based on validation performance (least mean absolute error). The final model with the optimal parameter set was then evaluated as the mean absolute error between model outputs and the testing set of 100 trials.

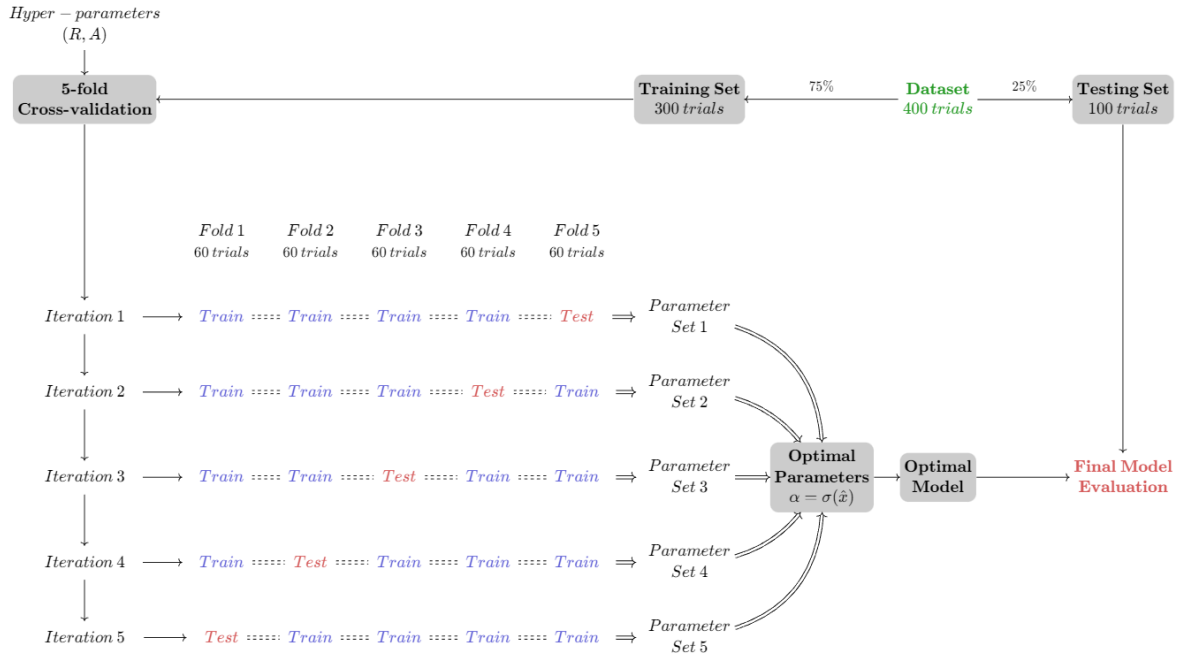

Figure 8: **Cross-validation overview:** Overview of model evaluation using a 5-fold cross-validation.

Chapter IV: Results

Data quality

From a total of 6400 trials presented to the 20 participants, 5984 trials (93%) were retained after accounting for data quality criteria. A total of 270 trials were removed from analysis due to poor visual tracking (based on tracking error or blinks), and a total of 146 trials were removed due to invalid (41 trials) or outlier (105 trials) response times. Outlier response times were determined from an 1.5IQR rule (see), where the acceptable response times across subjects was 199-264 ms as a lower bound and 507-572 ms as an upper bound; such response time cutoffs are consistent with previous studies on manual response times during smooth pursuit [Khan et al., 2010].

Trials were not screened nor excluded due to saccades, as the unpredictable motion induced several "catch-up" saccade events (and often very little pure smooth pursuit) due to the rapid changes in motion direction. Conversely, very few saccades occurred during predictable motion. Hence saccade data was deemed a feature of the motion tracking, in line with the hypothesis of research question #3, which suggested the ability to use eye movement acceleration to infer target motion predictability.

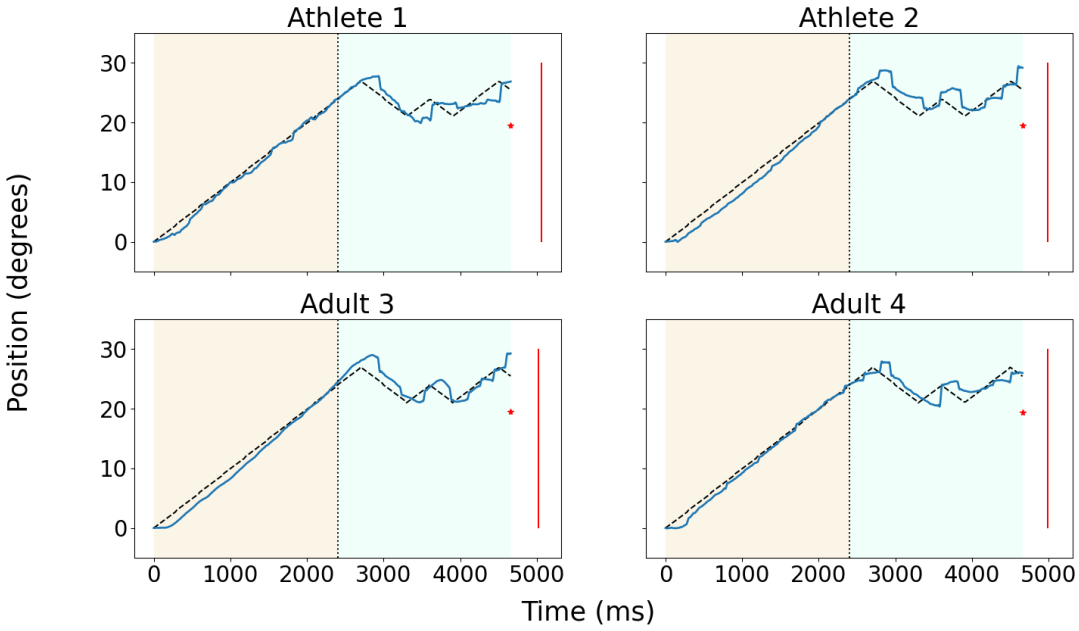

Figure 9: **Predictable-to-unpredictable motion trials:** Eye movement and response time data across subjects for predictable-to-unpredictable motion trials. Target (black dashes), eye (blue), and probe (red asterisk) position are represented, as well as moment of response time (red line) and the distinction of initial (yellow) and final (cyan) phases of motion.

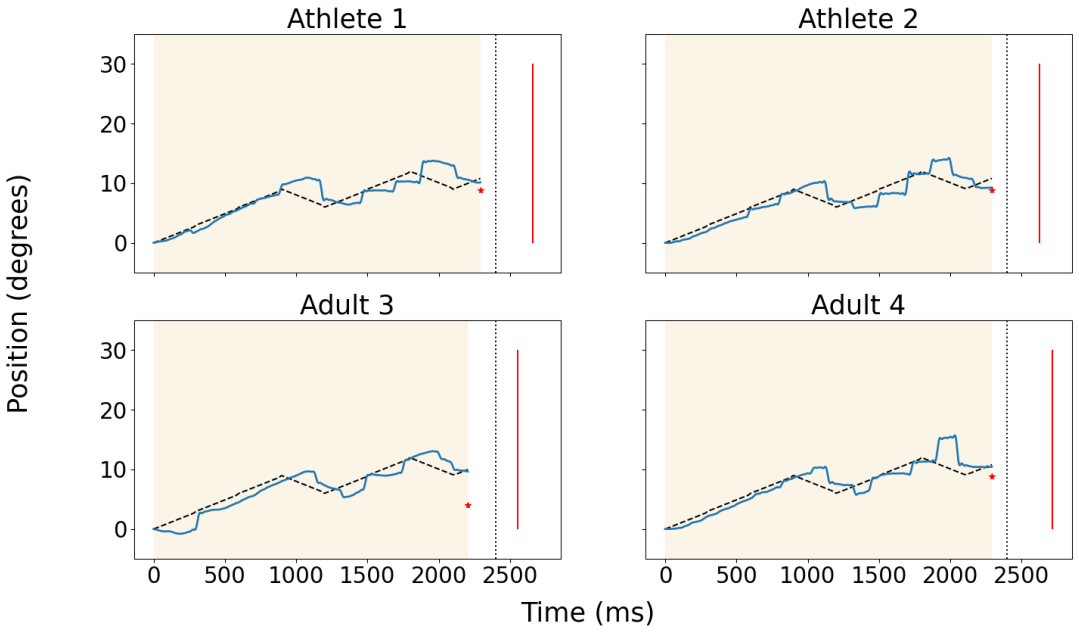

Figure 10: **Semi-predictable motion trials:** Eye movement and response time data across subjects for semi-predictable motion trials. Target (black dashes), eye (blue), and probe (red asterisk) position are represented, as well as moment of response time (red line) and the distinction of initial (yellow) and final (cyan) phases of motion.

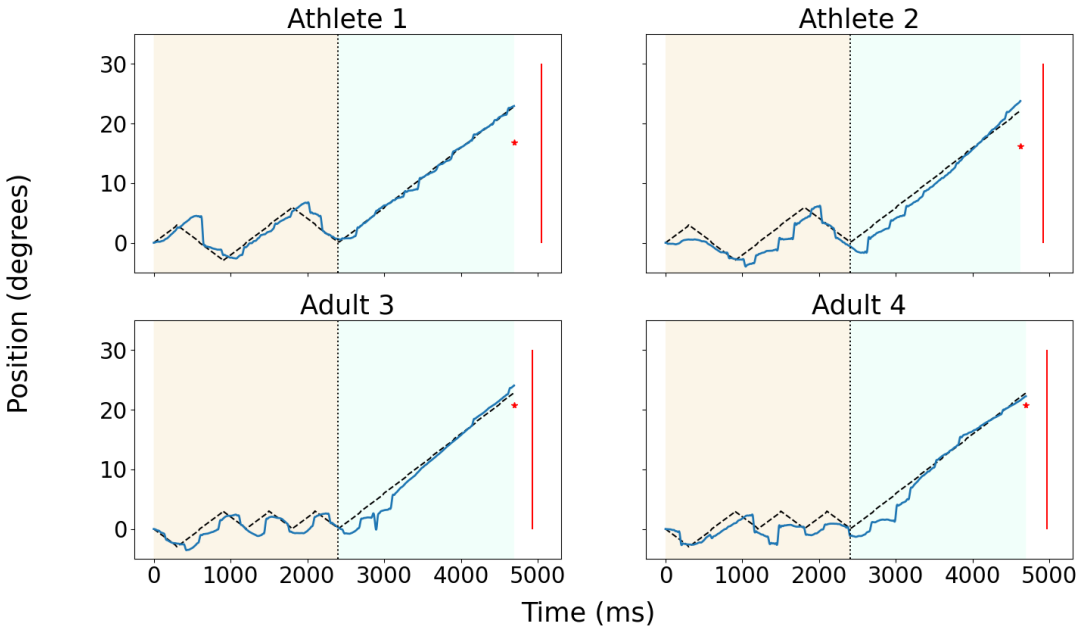

Figure 11: **Unpredictable-to-predictable motion trials:** Eye movement and response time data across subjects for unpredictable-to-predictable motion trials. Target (black dashes), eye (blue), and probe (red asterisk) position are represented, as well as moment of response time (red line) and the distinction of initial (yellow) and final (cyan) phases of motion.

Population characteristics

Visual acuity

A two-way mixed ANOVA was performed to analyze the effect of group (between factor) and visual test (within factor) on visual acuity (Figure 12). There was no statistically significant interaction between group and visual test factors ($F(2, 36) = 1.89$, $p = 0.17$).

An analysis of simple main effects showed that both the group factor had a statistically significant effect on visual acuity ($p = 0.008$), as well as the visual test factor ($p < 0.001$).

Group	n	Visual Assessment	Score Mean (logMAR)	Score SD (logMAR)
		Static VA	-0.162	0.033
Athletes	10	Linear DVA	0.034	0.081
		Random DVA	0.074	0.086
		Static VA	-0.136	0.072
Adults	10	Linear DVA	0.138	0.066
		Random DVA	0.136	0.075

Table 1: **Visual acuity scores:** Mean and standard deviations (SD) for visual acuity scores for static (SVA) and dynamic (DVA) assessments, by group.

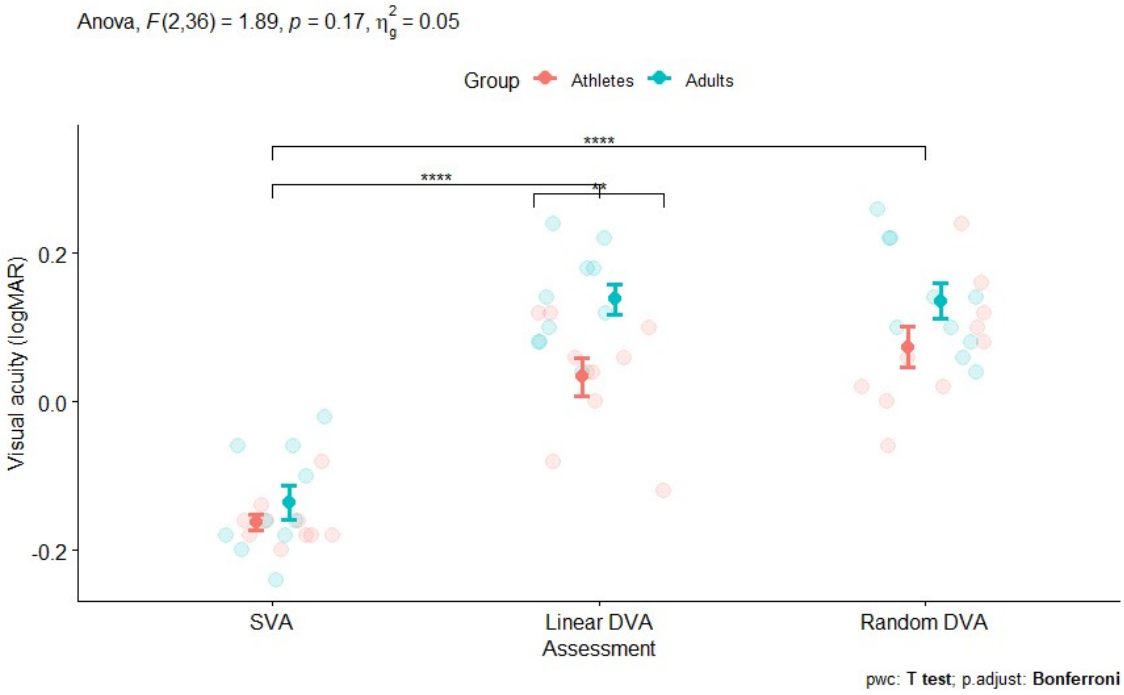

Figure 12: **Visual acuity scores:** Visual acuity for static (SVA) and dynamic (DVA) assessments. Asterisks (*) denote statistical significance level of $p < 0.05$.

Baseline response times

A two-way mixed ANOVA was performed to analyze the effect of group (between factor) and probe distance (within factor) on baseline manual response times from static trials (, Figure 13). There was no statistically significant interaction between group and probe distance (F(1, 18) = 0.00067, p = 0.98).

An analysis of simple main effects showed that the athlete had faster baseline response times than the adult group ($p < 0.01$), and that for both groups response times to probes

at 2° were faster than to probes at 6° ($p = 0.02$).

Group	n	Probe distance (°)	RT Mean (ms)	RT SD (ms)
Athletes	10	2	273	15
		6	295	15
Adults	10	2	301	28
		6	322	36

Table 2: **Baseline response times:** Mean and standard deviations (SD) for baseline manual response times (ms) for each probe distance (2°, 6°), by group.

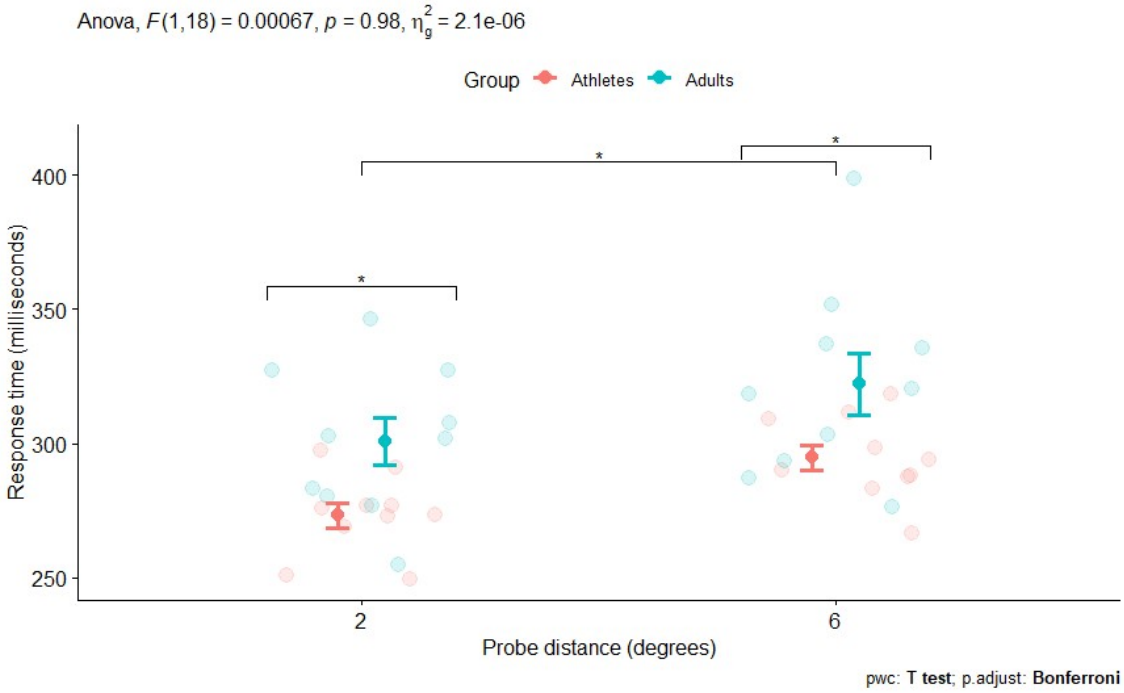

Figure 13: **Baseline response times:** Baseline manual response times for each group and distance. Asterisks (*) denote statistical significance level of $p < 0.05$.

Research question 1

Para-foveal normalized response times

A three-way mixed ANOVA was performed to analyze the effect of group (between factor), motion predictability (within factor), and probe location (within factor) on normalized manual response times from the first phase of motion ($t \leq 2400ms$) of dynamic trials () for trials where probes flashed at a distance of 2° eccentricity (Figure 14, Table 4). There was

no statistically significant three-way interaction between group, motion predictability and probe location ($F(2, 36) = 0.004$, $p = 0.996$). However, there was a significant interaction between movement predictability and probe location ($F(2, 36) = 30.903$, $p<0.01$). Simple main effects analysis showed that factors of predictability ($F(2,36) = 60.28$, $p<0.01$) and probe location ($F(1, 18) = 118.48$, $p<0.01$) had significant effects on response times.

Pairwise comparisons revealed significantly faster response times to probes ahead of target motion compared to behind target motion during predictable motion for athletes ($p=0.01$) and adults ($p=0.02$). Overall, response times between each motion predictability condition were significantly different ($p<0.01$), with lowest response times for predictable motion, highest response times for unpredictable motion, and intermediate response times for semi-predictable motion.

Effect	DFn	DFd	F	p	ges
Group	1	18	0.23	0.64	1.10E-02
Predictability	2	36	60.28	<0.01	3.04E-01
Location	1	18	118.48	<0.01	5.20E-02
Group : Predictability	2	26	0.04	0.96	2.05E-01
Group : Location	1	18	1.73	0.21	8.06E-04
Predictability : Location	2	36	30.90	<0.01	3.60E-02
Group : Predictability : Location	2	36	0.004	0.99	4.35E-06

Table 3: **Research question #1 ANOVA results for para-foveal probes:** Three-way mixed ANOVA results for the effect of group, motion predictability, and probe location on normalized response times (ms) for a probe distance of 2°.

Group	Motion	Probe location	n	NRT Mean (ms)	NRT SD (ms)
Athletes	$P_T = 0$	behind	10	80	30
		ahead	10	49	16
	$P_T = 0.25$	behind	10	98	19
		ahead	10	88	24
	$P_T = 0.5$	behind	10	117	30
		ahead	10	117	30
Adults	$P_T = 0$	behind	10	74	36
		ahead	10	40	26
	$P_T = 0.25$	behind	10	92	44
		ahead	10	79	43
	$P_T = 0.5$	behind	10	114	49
		ahead	10	109	50

Table 4: **Normalized para-foveal response times:** Mean and standard deviations (SD) for normalized response times (ms) to probes located 2° ahead or behind target motion that was predictable ($P_T = 0$), semi-predictable ($P_T = 0.25$), and unpredictable ($P_T = 0.5$), by group.

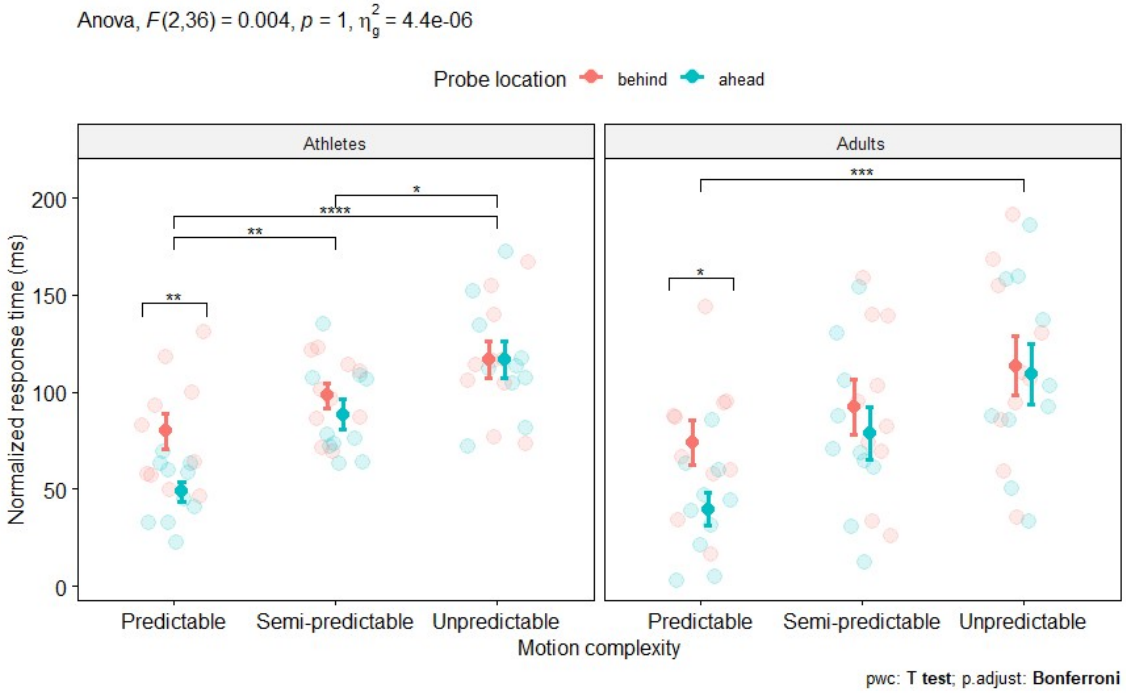

Figure 14: **Normalized para-foveal response times:** Normalized response times (ms) to probes located 2° ahead or behind target motion by group and motion predictability.

Peripheral normalized response times

A three-way mixed ANOVA was performed to analyze the effect of group (between factor), motion predictability (within factor), and probe location (within factor) on normalized manual response times from the first phase of motion ($t \leq 2400ms$) of dynamic trials () for trials where probes flashed at a distance of 6° eccentricity (Figure 15, Table 6). There was a statistically significant three-way interaction between group, motion predictability and probe location (F(2, 36) = 11.876, p<0.01). There were no significant two-way interactions. Simple main effects analysis showed that factors of predictability (F(2,36) = 18.96, p<0.01) and probe location (F(1, 18) = 10.15, p<0.01) had significant effects on response times.

Pairwise comparisons revealed significantly faster response times to probes ahead of target motion compared to behind target motion during predictable motion for only the adult group (p<0.01). Normalized response times during predictable motion were also significantly faster than during unpredictable motion for the athlete group (p<0.01).

Effect	DFn	DFd	F	p	ges
Group	1	18	0.11	0.74	0.005
Predictability	2	36	18.96	<0.01	0.137
Location	1	18	10.15	<0.01	0.008
Group : Predictability	2	36	0.32	0.73	0.003
Group : Location	1	18	2.72	0.12	0.002
Predictability : Location	2	36	2.42	0.10	0.001
Group : Predictability : Location	2	36	11.88	<0.01	0.006

Table 5: **Research question #1 ANOVA results for peripheral probes:** Three-way mixed ANOVA results for the effect of group, motion predictability, and probe location on normalized response times (ms) for a probe distance of 6°.

Group	Motion	Probe location	n	NRT Mean (ms)	NRT SD (ms)
Athletes	$P_T = 0$	behind	10	59	27
		ahead	10	60	24
	$P_T = 0.25$	behind	10	87	39
		ahead	10	83	36
	$P_T = 0.5$	behind	10	111	47
		ahead	10	103	52
Adults	$P_T = 0$	behind	10	72	44
		ahead	10	48	36
	$P_T = 0.25$	behind	10	81	58
		ahead	10	72	53
	$P_T = 0.5$	behind	10	98	56
		ahead	10	96	58

Table 6: **Normalized peripheral response times:** Mean and standard deviations (SD) for normalized response times (ms) to probes located 6° ahead or behind target motion that was predictable ($P_T = 0$), semi-predictable ($P_T = 0.25$), and unpredictable ($P_T = 0.5$), by group.

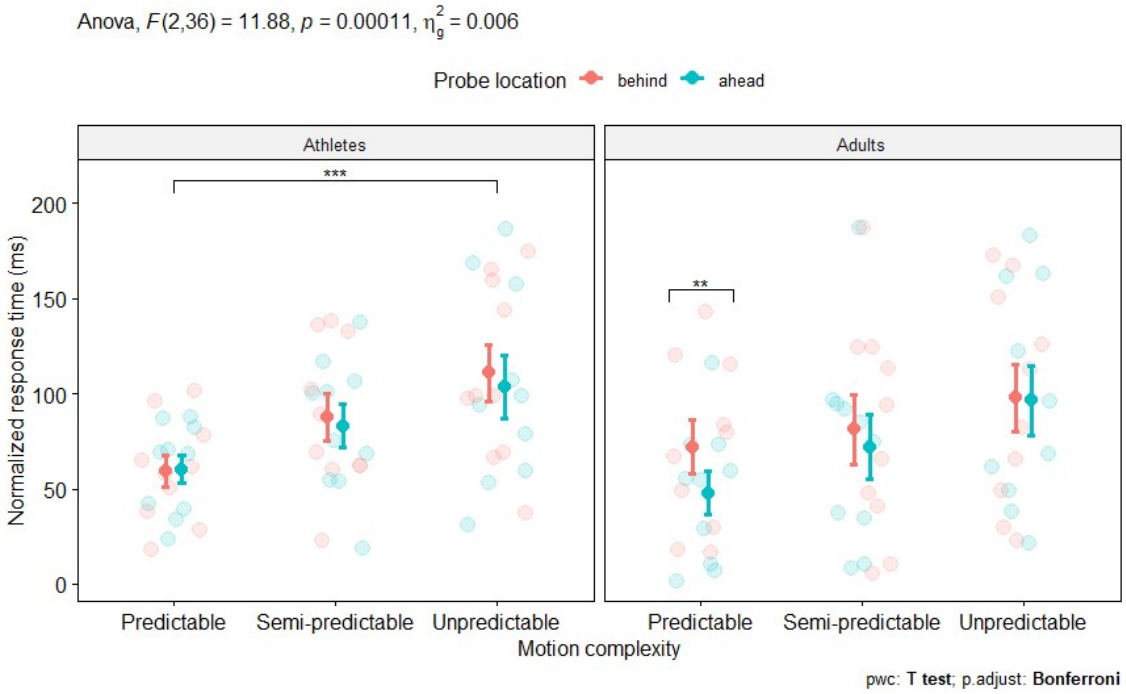

Figure 15: **Normalized peripheral response times:** Normalized response times (ms) to probes located 6° ahead or behind target motion by group and motion predictability.

Visual tracking error

A two-way mixed ANOVA was performed to analyze the effect of group (between factor) and motion predictability (within factor) on visual tracking error, as defined in (Figure 16). There was no statistically significant interaction between group and motion predictability factors ($F(2, 36) = 0.485$, $p = 0.62$).

An analysis of simple main effects showed that motion predictability had a statistically significant effect on visual tracking error ($F(2, 36) = 183.559$, $p<0.01$). Pairwise comparisons revealed the tracking error was significantly different ($p<0.01$) across all motion predictability levels, with predictable motion resulting in far lower error ($p<0.0001$) compared to semi-predictable and unpredictable motion, and unpredictable motion resulting in less error than semi-predictable motion.

Group	Motion Predictability	n	Mean error (degree)	SD error (degree)
Athletes	$P_T = 0$	10	0.68	0.12
	$P_T = 0.25$	10	1.07	0.08
	$P_T = 0.5$	10	1.01	0.12
Adults	$P_T = 0$	10	0.69	0.14
	$P_T = 0.25$	10	1.13	0.15
	$P_T = 0.5$	10	1.03	0.14

Table 7: **Visual tracking error:** Mean and standard deviations (SD) for visual tracking error (degrees) during predictable ($P_T = 0$), semi-predictable ($P_T = 0.25$), and unpredictable ($P_T = 0.5$) motion, by group.

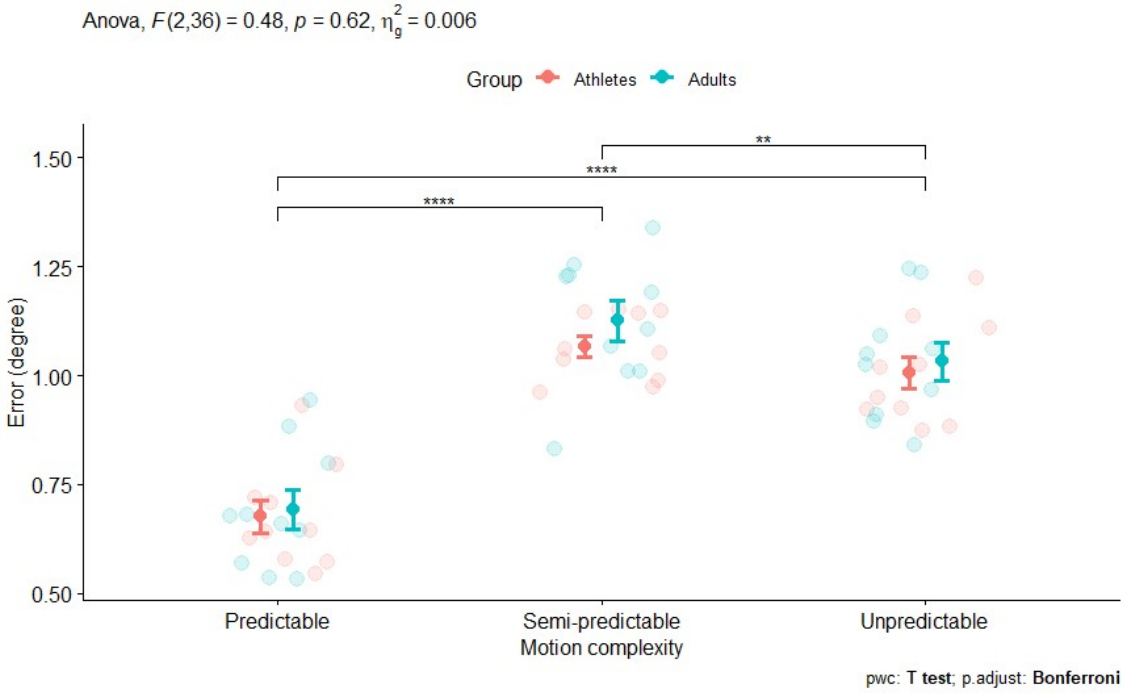

Figure 16: **Visual tracking error:** Visual tracking error (degree) by group and motion predictability.

Research question 2

Para-foveal response time bias

A two-way mixed ANOVA was performed to analyze the effect of group (between factor) and probe time window (within factor) on normalized response time bias () to probes flashed 2° about predictable-to-unpredictable target motion (Figure 17). There was no statistically significant interaction between group and probe time factors ($F(3, 54) = 0.541$, $p = 0.66$).

An analysis of simple main effects showed that probe time had a statistically significant effect on response time bias ($F(3, 54) = 14.769$, $p<0.01$). Pairwise comparisons revealed that the magnitude of response time bias was significantly less ($p<0.001$) in both of the first two time windows (1-1200 ms & 1201-2400 ms) which consisted of predictable motion, compared to both of the final two time windows (2401-3600 ms & 3601-4800 ms) which consisted of unpredictable motion.

A second two-way mixed ANOVA was performed to analyze the effect of group (be-

tween factor) and probe time window (within factor) on normalized response time bias to probes flashed 2° about unpredictable-to-predictable target motion (Figure 18). There was no statistically significant interaction between group and probe time factors (F(1.89, 34.04) = 1.74, p = 0.19).

An analysis of simple main effects showed that probe time had a statistically significant effect on response time bias (F(1.89, 34.04) = 5.52, p<0.01). Pairwise comparisons revealed that the magnitude of response time bias was significantly less (p<0.01) in both of the first two time windows (1-1200 ms & 1201-2400 ms) which consisted of unpredictable motion, compared to the final time window (3601-4800 ms) of predictable motion.

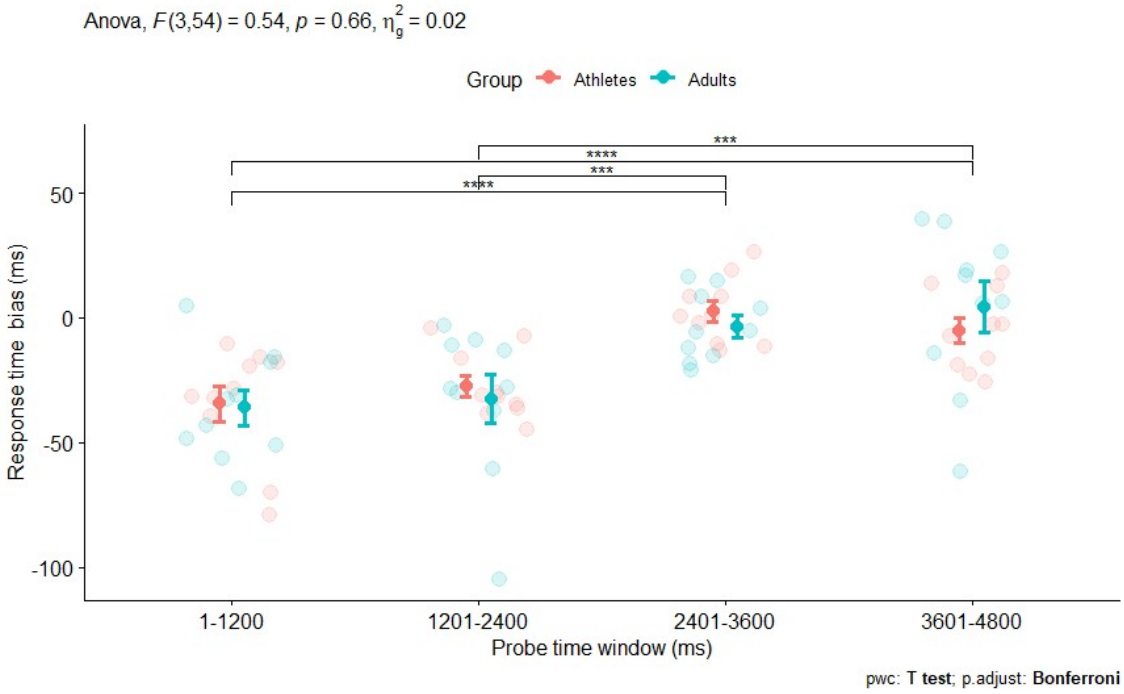

Figure 17: **Para-foveal RT bias to $P_T = 0 \rightarrow 0.5$ motion:** Response time bias to probes flashed 2° about predictable-to-unpredictable target motion, by probe time and group.

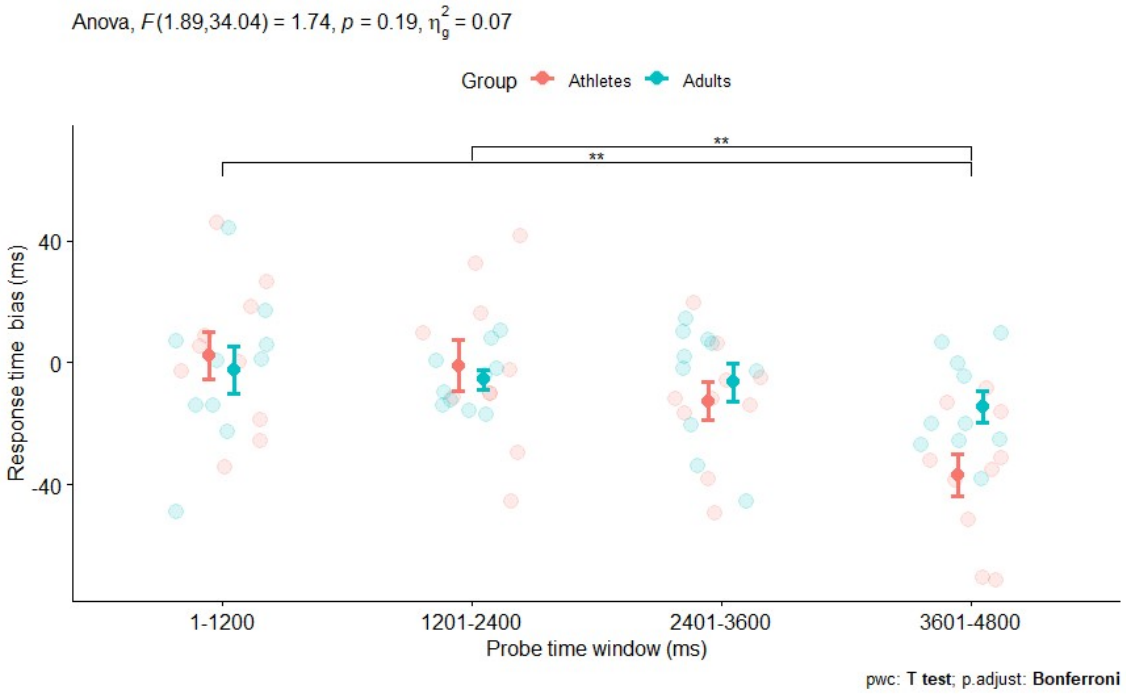

Figure 18: **Para-foveal RT bias to $P_T = 0.5 \rightarrow 0$ motion:** Response time bias to probes flashed 2° about unpredictable-to-predictable target motion, by probe time and group.

Peripheral response time bias

A two-way mixed ANOVA was performed to analyze the effect of group (between factor) and probe time window (within factor) on normalized response time bias to probes flashed 6° about predictable-to-unpredictable target motion (Figure 19). There was a statistically significant interaction between group and probe time factors (F(1.93, 34.75) = 3.657, p = 0.038).

Pairwise comparisons revealed a significantly larger magnitude of bias ($p<0.05$) in the adult group compared to the athlete group during both of the first two time windows (1-1200 ms & 1201-2400 ms) which consisted of predictable motion.

A second two-way mixed ANOVA was performed to analyze the effect of group (between factor) and probe time window (within factor) on normalized response time bias to probes flashed 6° about unpredictable-to-predictable target motion (Figure 20). There was no statistically significant interaction between group and probe time factors (F(3, 54) = 1.92, p = 0.14). There was no significant main effects from either group or probe

time factors.

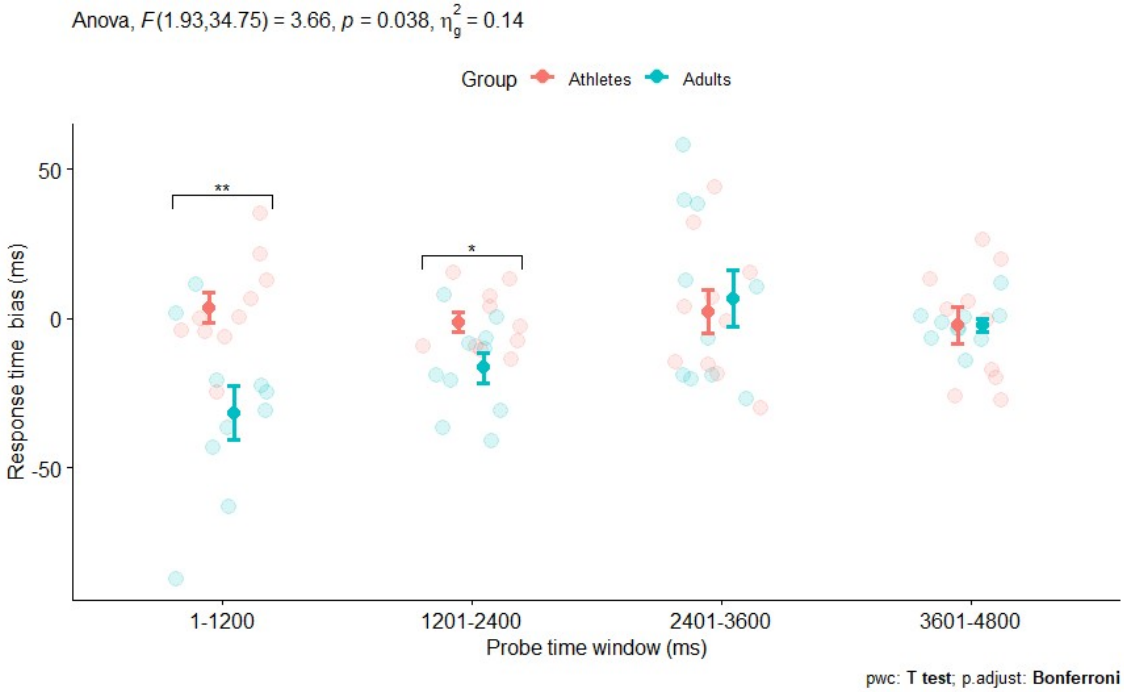

Figure 19: **Peripheral RT bias to $P_T = 0 \rightarrow 0.5$ motion:** Response time bias to probes flashed 6° about predictable-to-unpredictable target motion, by probe time and group.

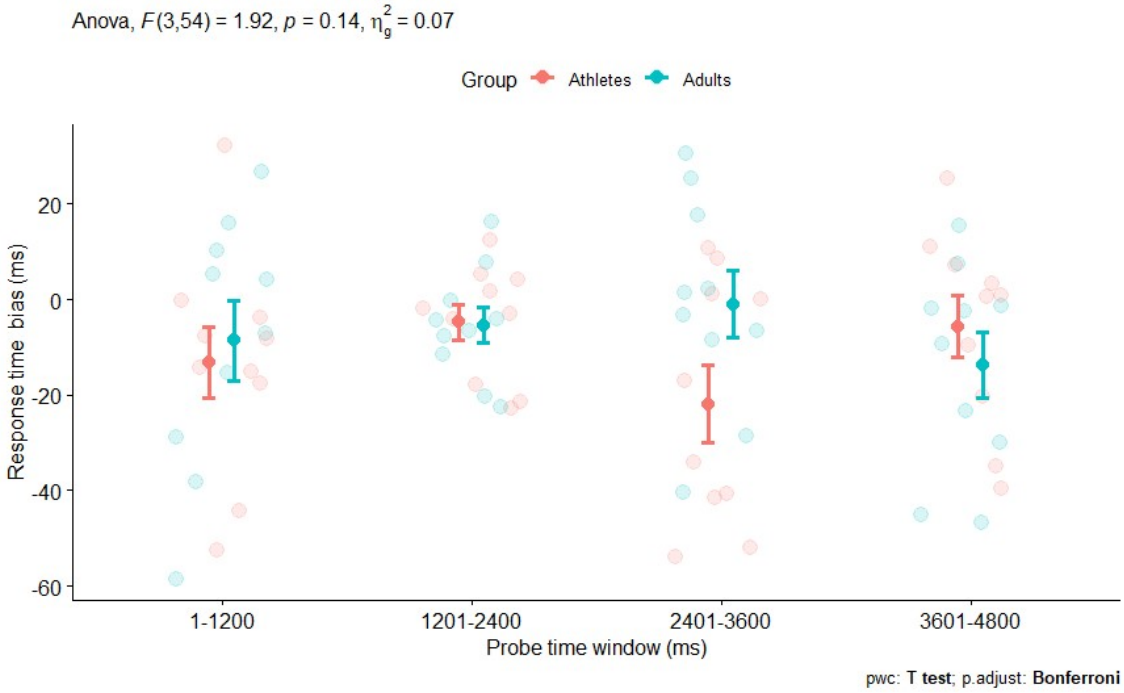

Figure 20: **Peripheral RT bias to $P_T = 0.5 \rightarrow 0$ motion:** Response time bias to probes flashed 6° about unpredictable-to-predictable target motion, by probe time and group.

Visual tracking error

A three-way mixed ANOVA was performed to analyze the effect of group (between factor), motion type (within factor), and motion phase (within factor) on visual tracking error of dynamic trials, as defined in (Figure 21). There was no statistically significant three-way interaction between group, motion type and motion phase ($F(1, 18) = 0.366$, $p=0.553$). There was a significant two-way interaction between motion type and motion phase ($F(1, 18) = 68.477$, $p<0.01$).

Pairwise comparisons revealed that for both athletes and adults there was a significant ($p<0.0001$) increase in visual tracking error across predictable-to-unpredictable motion phases, compared to no significant difference in tracking error across unpredictable-to-predictable motion phases.

Group	Motion type	Phase	n	Mean error (degree)	SD error (degree)
Athletes	Predictable-to-unpredictable	1	40	0.676	0.118
		2	40	0.989	0.174
	Unpredictable-to-predictable	1	40	1.006	0.118
		2	40	1.026	0.189
Adults	Predictable-to-unpredictable	1	40	0.693	0.141
		2	40	1.014	0.167
	Unpredictable-to-predictable	1	40	1.032	0.136
		2	40	1.101	0.259

Table 8: **Visual tracking error across motion phases:** Mean and standard deviations (SD) for visual tracking error (degrees) by motion type, motion phase, and group.

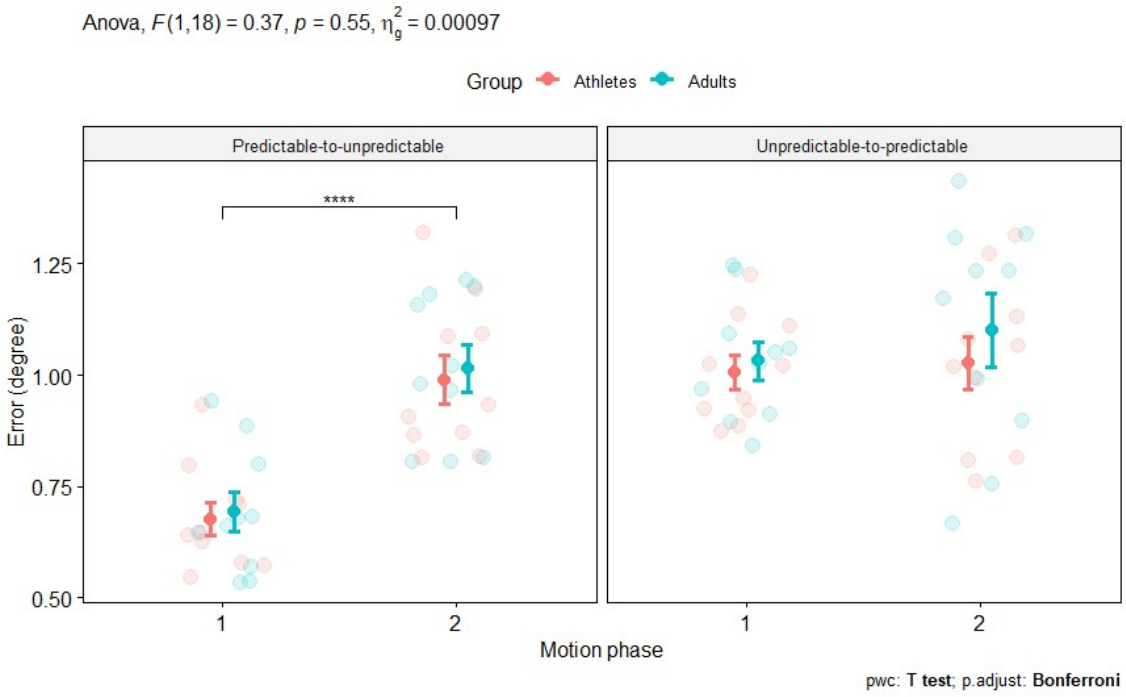

Figure 21: **Visual tracking error across motion phases:** Visual tracking error based on motion type (predictable-to-unpredictable or unpredictable-to-predictable), motion phase (1 or 2) and group (athlete or adult). For the predictable-to-unpredictable motion type, motion phase 1 was predictable and motion phase 2 was unpredictable. For the unpredictable-to-predictable motion type, motion phase 1 was unpredictable and motion phase 2 was predictable.

Eye movement dynamics

A three-way mixed ANOVA was performed to analyze the effect of group (between factor), motion type (within factor), and motion phase (within factor) on the number of saccades

per motion step (300 ms) of dynamic trials (Figure 22). While there was no statistically significant three-way interaction between group, motion type and motion phase ($F(1, 18) = 0.425$, $p=0.523$), there was a significant two-way interaction between motion type and motion phase ($F(1, 18) = 23.296$, $p<0.01$).

Another three-way mixed ANOVA was performed to analyze the effect of group (between factor), motion type (within factor), and motion phase (within factor) on the average saccade amplitude per motion step (300 ms) of dynamic trials (Figure 23). Again, there was no statistically significant three-way interaction between group, motion type and motion phase ($F(1, 18) = 0.297$, $p=0.592$), and there was a significant two-way interaction between motion type and motion phase ($F(1, 18) = 13.162$, $p<0.01$).

Pairwise comparisons revealed that for both athletes and adults there was a significant ($p<0.0001$) increase in the number of saccades and average saccade amplitude per motion step across predictable-to-unpredictable motion phases, compared to no significant difference across unpredictable-to-predictable motion phases.

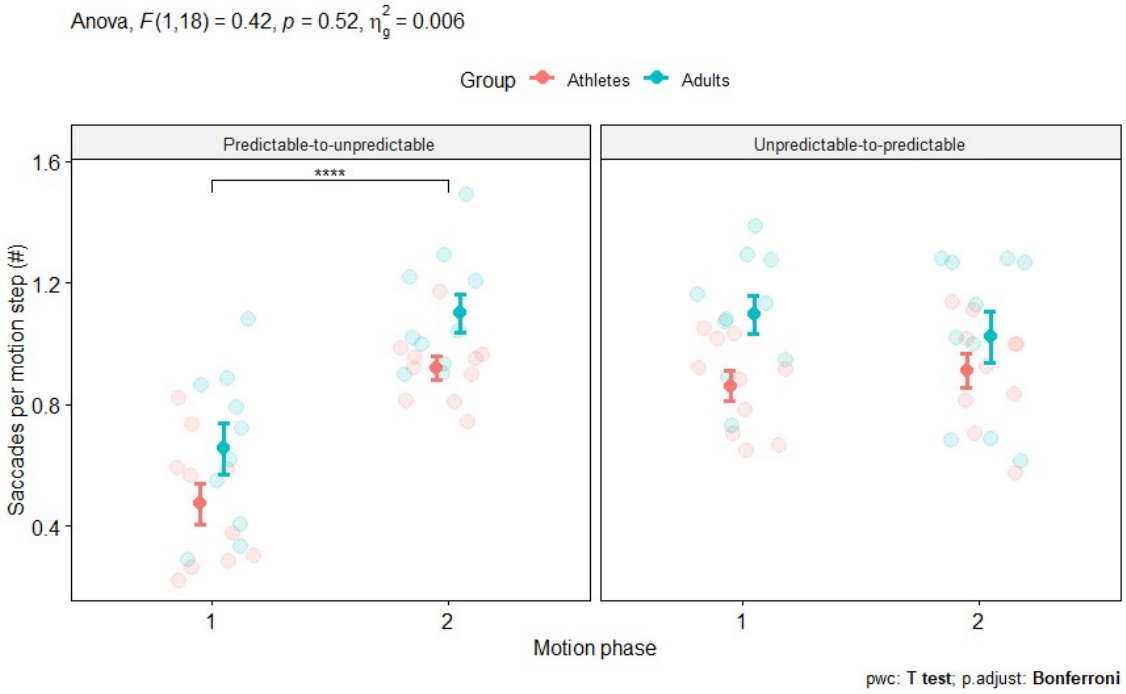

Figure 22: **Saccades per motion step:** Average number of saccades per motion step (300 ms) based on motion type (predictable-to-unpredictable or unpredictable-to-predictable), motion phase (1 or 2) and group (athlete or adult). For the predictable-to-unpredictable motion type, motion phase 1 was predictable and motion phase 2 was unpredictable. For the unpredictable-to-predictable motion type, motion phase 1 was unpredictable and motion phase 2 was predictable.

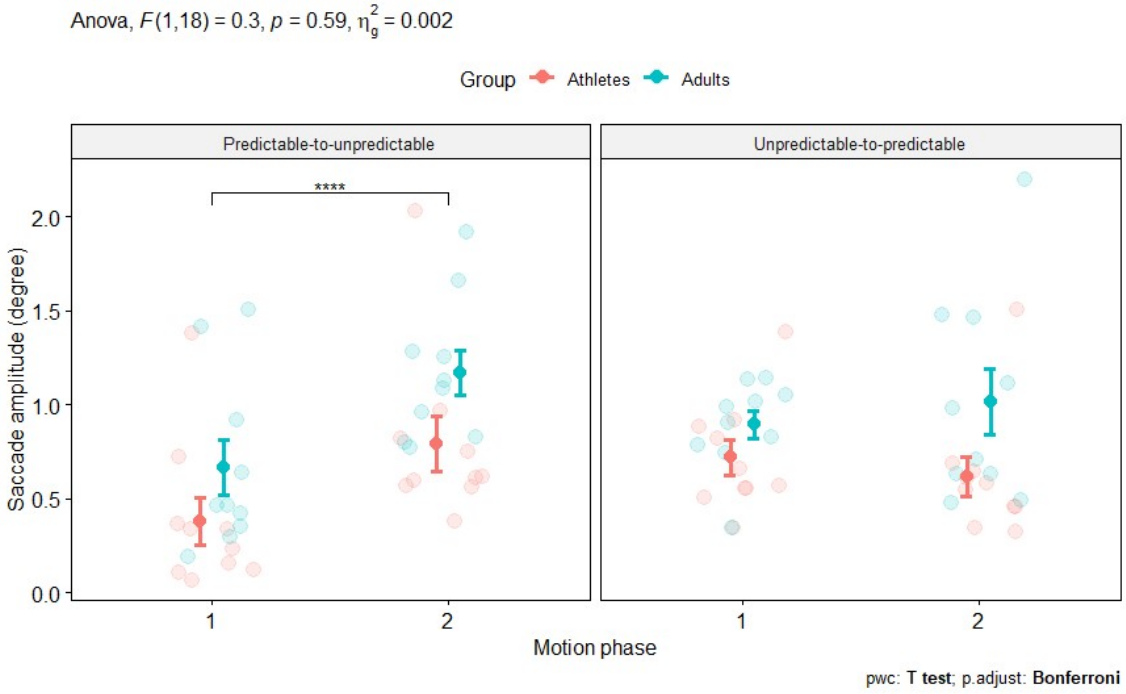

Figure 23: **Saccade amplitudes per motion step:** Average saccade amplitude per motion step (300 ms) based on motion type (predictable-to-unpredictable or unpredictable-to-predictable), motion phase (1 or 2) and group (athlete or adult). For the predictable-to-unpredictable motion type, motion phase 1 was predictable and motion phase 2 was unpredictable. For the unpredictable-to-predictable motion type, motion phase 1 was unpredictable and motion phase 2 was predictable.

Research question 3

Model hyper-parameters

The diagonals of the measurement covariance matrix (R) for the Kalman filter component of the model was tuned by a linear grid search, by testing the mean absolute error between filter outputs and experimental eye position data for R in the range of 0 to 5, in steps of 0.04. For reference, in previous implementations of the Kalman filter for smooth pursuit simulations, the equivalent value of R was set at 5 degrees^2 [de Xivry et al., 2013]. The optimal value of R was determined as 3.52 degrees^2 (Figure 24). Acceleration values were rescaled using a sigmoid function centered at 63.75 $^{\circ}/_{sec^2}$, which was the 3^{rd}-quartile of simulated pursuit acceleration values.

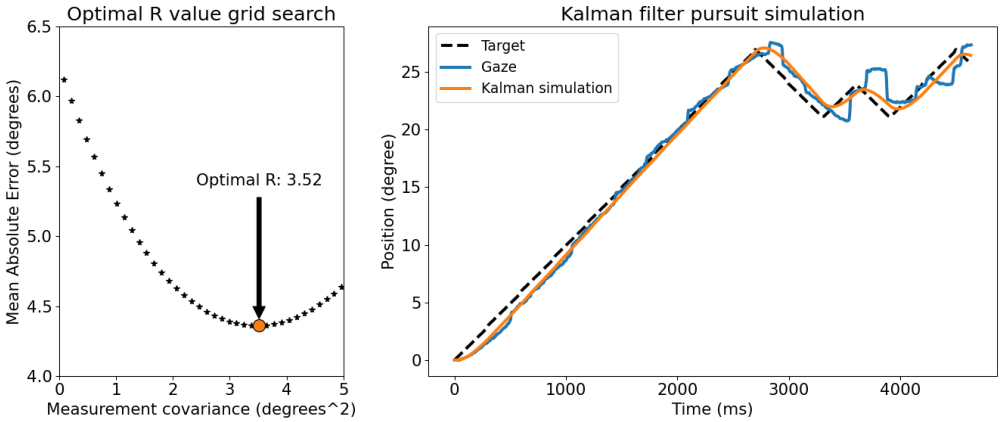

Figure 24: **Kalman filter hyperparameter selection: Left**: Mean absolute error of pursuit simulation compared to experimental eye position, based on Kalman filter measurement covariance (R); **Right**: Demonstration of Kalman filter pursuit simulation compared to experimental eye tracking data.

Model validation

A 5-fold cross-validation scheme was used to find the optimal parameters of the α-scaling sigmoid function, specifically to determine a bias rate parameter (k). Model outputs were compared to normalized experimental response time bias data, and mean absolute error was used as a metric to measure performance (Table 9). Based on cross-validation results, a logistic rate of 54 (taken from the first iteration of the cross-validation) was used to define the α function of the leaky integrator component.

Iteration	n	Logistic k	Validation MAE (%)	Validation MAE SD (%)
1	60	58	19.4	5.5
2	60	54	18.5	5.1
3	60	52	19.8	5.3
4	60	56	18.8	4.3
5	60	52	19.6	4.6

Table 9: **Cross-validation results:** Table of 5-fold cross-validation results for the selection of logistic k parameter to tune the sigmoid of α function of the model leaky integrator component.

Model performance

The final model was evaluated against a test set of 100 trials separate from the 300 trials used in cross-validation (Figure 25). The mean absolute error between model and experimental bias values was 18.6% (SD=0.04%).

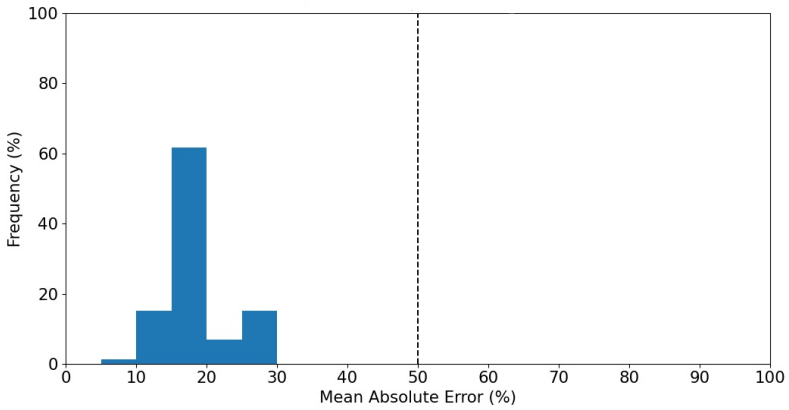

Figure 25: **Model test performance:** Model performance against the test set (n=100), measured by mean absolute error between normalized experimental response time bias values and model bias estimation.

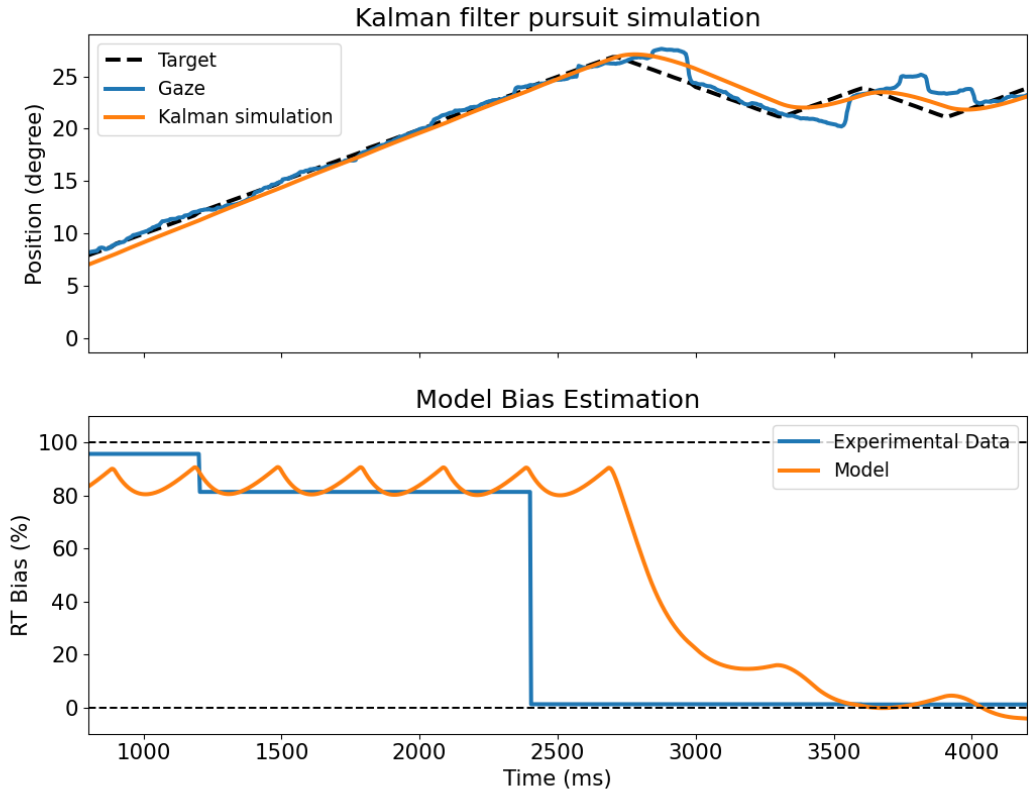

Figure 26: **Model performance for predictable-to-unpredictable motion:** Model bias estimation performance against experimental data for predictable-to-unpredictable motion. **Top**: Pursuit simulation against experimental position data; **Bottom**: Bias estimation against experimental bias data.

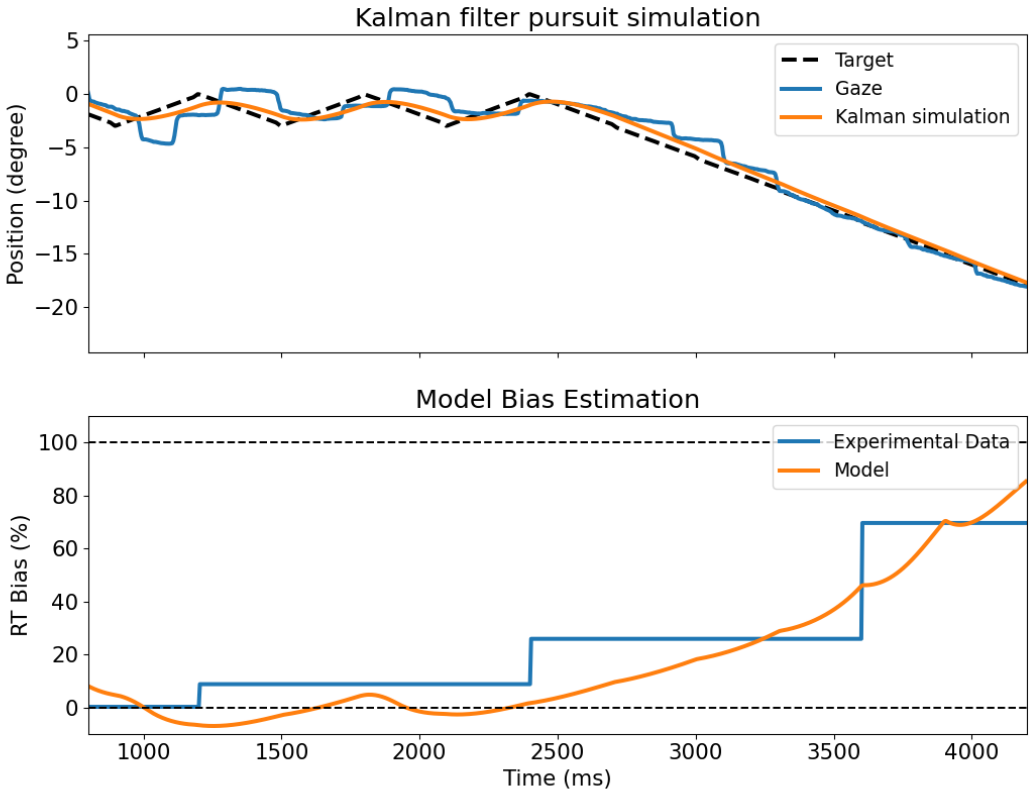

Figure 27: **Model performance for unpredictable-to-predictable motion:** Model bias estimation performance against experimental data for unpredictable-to-predictable motion. **Top**: Pursuit simulation against experimental position data; **Bottom**: Bias estimation against experimental bias data.

Chapter V: Discussion

Comparable performance between athletes and adults

The baseline response time and dynamic visual acuity scores indicated a significance difference between athlete and adult groups. These results were in line with previous studies demonstrating faster simple visual reaction times [Ando et al., 2001] and better dynamic visual acuity scores [Yee, 2017] in athletes compared to healthy controls. However, despite the clear distinction in baseline ability, for the most part there was a lack of significant difference between groups in terms of normalized response time performance during motion tracking. One significant difference between groups was that in the athlete group a bias in response time during predictable motion tracking was only present at a probe distance of 2°, while in the healthy controls the bias was present even for probes at a distance of 6° (Figure 15). The discrepancy in the spread of attention ahead of target motion may indicate a higher level of processing efficiency in athletes, by only selectively attending to where the target is likely to be next, rather than broadly in the direction of target motion. It may also be that team-sport athletes, who are often tracking multiple objects in the visual field, may need such discrimination of responses to foveal and peripheral stimuli so that overtly tracking a direct opponent does not affect the ability to covertly track additional opponents in the peripheral visual field. Overall, the comparable performance between athletes and controls may be due to the lack of a sport-specific motion stimulus and constrained experimental setup, potentially reducing any advantage in the trained behaviours of the athletes, which are often practiced in more complex, open environments.

Visual attention and motion predictability

The primary objective of this work was to determine whether the bias in visual attention ahead of target motion during visual motion tracking was dependent on the predictability of target motion. Previously, studies had focused assessing visual attention during pure smooth pursuit of targets moving in a predictable, linear fashion

[Lovejoy et al., 2009, Khan et al., 2010]. However, smooth pursuit of unpredictable motion is a near paradoxical scenario, as unpredictability requires discontinuity to some extent (i.e. a lack of smoothness). As such, the current work investigated visual motion tracking in general, as a combination of pursuit and ("catch-up") saccadic eye movements. Regardless, both smooth pursuit and saccades have demonstrated a bias towards the anticipated future location of a target [Khan et al., 2010, Harrison et al., 2012], and as such, the question of the role of predictability in that future target location on attentional bias during a dynamic motion tracking task remains relevant.

In analyzing the response times to probes flashed ahead versus behind target motion during predictable, semi-predictable, and unpredictable tracking conditions, only the predictable condition demonstrated any significant bias towards faster response times ahead of target motion (Figures 14, 15). These results align with previous studies demonstrating a bias during pursuit of predictable motion [Donkelaar and Drew, 2002, Khan et al., 2010]. Additionally, the bias in response to predictable motion was found to take longer to accumulate (post-random motion) than it took to de-accumulate when followed by random motion (Figures 17, 18).

Motion predictability, eye movements, & covert attention

An asymmetrical response to motion predictability was not only apparent in response time data, but in eye movement quality as well. Unpredictable motion had a larger number of saccadic eye movements, as well as larger visual tracking error (Figure 16, 21). Additionally, while very few saccades were present when the initial motion phase consisted of predictable motion, the same predictable motion had significantly more catch-up saccades of larger amplitudes when it followed random motion (Figures 22, 23).

The occurrence of catch-up saccades have been shown to be triggered by the *predicted* future eye position error [Nachmani et al., 2020]. The emphasis on the role of predicted error, as opposed to current error, provides some explanation as to why predictable motion induced varying amounts of catch-up saccades depending on whether it was the preceded or followed an unpredictable phase of motion. If the trigger was simply current error, such

an asymmetry should not be apparent. However, if the trigger is the current prediction of future error, then when the current motion phase is random then the future prediction of error will be high (Figure 21) and thus induce catch-up saccades. Furthermore, when transitioning from unpredictable to predictable motion, initially the prediction of future error will still be high and as such, catch-up saccades will continue to be triggered during predictable motion (Figure 22).

From an examination of raw data traces, the scarcity of saccades in the initial predictable motion condition (Figure 9) indicates that there was a latent effect of unpredictable motion affecting the quality of eye movements into any subsequent motion phases. Finally, the frequency and amplitude of saccades decreased as more time was spent tracking predictable motion which could be reflective of the possible adaptation of future position error predictions (Figure 11).

Dynamics of attentional bias

Prediction is a requirement to visually track even the simplest of motion without any lag due to the time needed for information to move through the nervous system [Brouwer et al., 2002]. Visual attention to a portion of the visual field results in greater allocation of cognitive resources to process that area [Müller et al., 2003], and as such could facilitate the faster or higher-resolution (or both) visual processing of stimuli in the attended area [Eriksen and Colegate, 1970, Carrasco and McElree, 2001]. As such, a lack of prior information, or unreliable information, of where a stimulus may appear implies that pre-allocating any resources to process any one area of the visual field provides no particular advantage than any other area. In fact, should the stimulus appear in a location other than the one currently being attended to, the time needed to de-allocate and re-allocate cognitive resources to the new location would be dis-advantageous in terms of processing delays and energy consumption [Posner et al., 1984]. In analyzing response times to probes ahead versus behind target motion, a bias in response times only occurred in the predictable condition and not in semi- and unpredictable conditions, despite the evenly spaced levels of motion predictability (Figure 14). That response time bias was not also

evenly spaced indicates a conservative attentional allocation scheme to avoid a possible temporal penalty for incorrectly attending to a location. This is further supported by results demonstrating that the accumulation of bias ahead of target motion was gradual compared to the sudden de-accumulation of bias (Figure 17, 18).

Neuro-physiological underpinnings of attentional bias

Visual attention, motion prediction, and eye movements span several areas of the nervous system, and as such specifying a single region responsible for distribution of attention proves difficult. In particular, the posterior parietal cortex [Steinmetz and Constantinidis, 1995], the frontal eye fields [Schall, 2004], and the lateral intra-parietal cortex [Bisley and Goldberg, 2003] have been shown to be involved in attention re-mapping during the planning of eye movements. The frontal eye fields (FEF) are also involved in motion estimation, especially in the absence (transient disappearance) of visual information [Barborica and Ferrera, 2003]. The FEF may also interact with the middle temporal region (MT) as a possible location for prior information of pursuit velocity which is needed for predicting motion dynamics and initiating pursuit, although the supplementary eye fields and lateral intra-parietal cortex may also be candidates for this interaction [Yang et al., 2012]. As such, the FEF may act as the location for the update step in the Kalman filtering scheme of the attention model (Figure 5), where measurements from MT are used to adjust motion velocity priors based on stored velocity priors.

The representation of target motion dynamics (i.e. an "internal model" of the tracked object) plays a large role in motion prediction, and the cerebellum contributes significantly towards housing and updating such an internal representation. The medium for the representation may be in the form of tonic spiking of Purkinje cells, which not only have been shown to encode velocities for arm or eye movements [Coltz et al., 1999], but also for the movements of external objects [Cerminara et al., 2009]. The cerebellum may then act as a prediction step in a Kalman filtering scheme, by applying the stored internal model of object motion to predict future motion steps.

While the update and prediction steps of a Kalman filter scheme, potentially located

in the FEF and cerebellum may not be directly connected anatomically, there may be an indirect connection via the superior colliculus (SC). The SC works in conjunction with the vermis in the cerebellum to integrate the retinal slip and predicted error signals required for the triggering of saccades [Schreiber et al., 2006, Keller et al., 1996]. As such, the eye movement circuit among the superior colliculus, vermis, & paramedian pontine nuclei for dynamically gating smooth pursuit and saccade eye movements [Krauzlis, 2004] may also be the location of a circuit for the "leaky integration" of the predicted eye position error signals for the modulation of attentional bias.

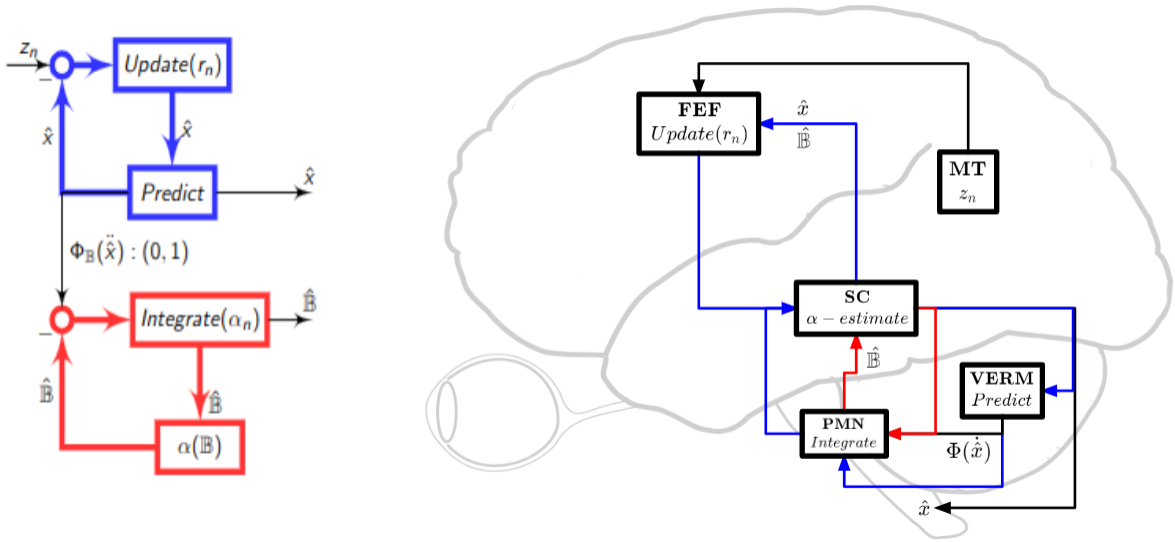

Figure 28: **Neuro-phsyiological implementation of computational framework:** Computational scheme of attentional modulation from state-space model used in research question #3 (**left**) and potential neuro-physiological implementation (**right**). MT: middle temporal lobe, FEF: frontal eye fields, SC: superior colliculus, VERM: vermis, PMN: paramedian pontine nuclei.

The manner in which a magnitude of attentional bias translates into visual (and subsequently, manual) response times might be explained using a linear approach to threshold with ergodic rate (LATER) model [Noorani and Carpenter, 2016]. According to the LATER model, the response to a stimulus is dependant on two parameters: a baseline activation level for the response decision signal and the rate of evidence accumulation. Once enough evidence is accumulated to activate the decision signal past a threshold, a response is produced (Figure 29). While the model has been used to ex-

plain various psycho-physical behaviours, a neuro-physiological correlates have also been found at the neuronal level in the visual pathway, specifically in the frontal eye fields [Hanes and Schall, 1996]. Recordings from neurons in the FEF of rhesus monkeys revealed that the variability in saccadic response times was directly correlated with the rate of accumulation in neural activity to the point of threshold, resulting in a saccade to the target stimulus. Accordingly, the bias value obtained from "leaky integration" in the state-space model would then be transmitted back from the superior colliculus to modulate the distribution of the activation rates of neurons in the frontal eye fields, and hence drive the bias in response times (Figure 30).

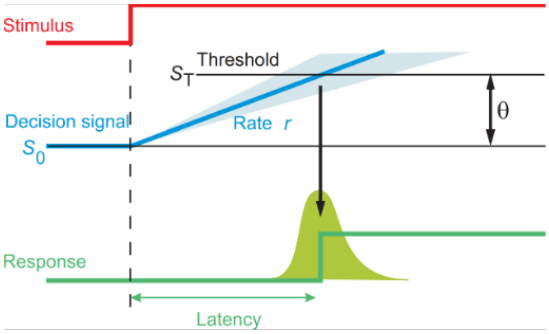

Figure 29: **LATER model:** LATER model of decision responses, from [Noorani and Carpenter, 2016]. Given a fixed starting activation level, evidence accumulates at a stochastic rate resulting in variability in response time latencies.

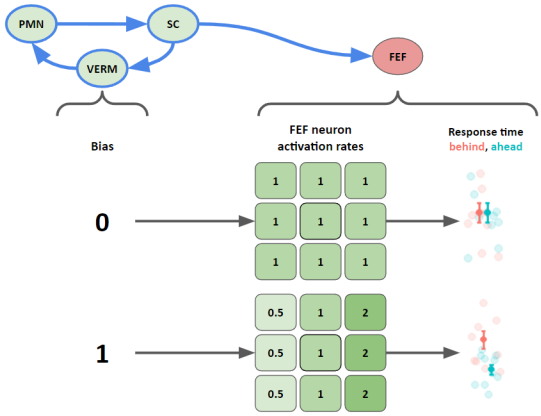

Figure 30: **Neuro-physiological implementation of bias**: A possible feedback pathway from the superior colliculus to the frontal eye fields could transform integrated bias values (Figure 28) into the appropriate distribution of neural activation rates to generate response time bias.

Chapter VI: Conclusion

Study limitations

Regarding the stimulus motion, the simple, random-walk motion generation model () which was constrained to the horizontal axis is not representative of the complex, multi-dimensional motion observed in the real-world. Furthermore, there were only two types of multi-phase motion used: predictable-to-unpredictable and unpredictable-to-predictable. The use of intermediate phases, or more continuous motion generation methods, may help to simulate motion resembling ecological motion behaviours. The lack of natural motion, or any other contextual factors (such as a ball target avatar, or sport field background), improved the experimental control regarding visual factors such as luminance and reference lines, as well as limiting any visual behaviours associated with sport specific contexts. Conversely, the added control may have been a factor in the resulting lack of significance between athletes and adults regarding motion tracking and response time bias data.

Regarding the probing of attention, the assumption was made that manual response time to a salient probe reflects visual spatial attention, as indicated by previous studies [Posner et al., 1980, Donkelaar and Drew, 2002, Khan et al., 2010]. Attention response probes were limited to ahead or behind motion (i.e. along the line of motion) and only at two distances (2° and 6°), due to the need to limit the number of trials per collection session to avoid participant fatigue. By comparison, previous studies investigating smooth pursuit and attention bias probed in all directions around target motion at a range of 1° to 5° [Khan et al., 2010], however these were collected over several sessions as opposed to a single collection session for the current study. Additionally, probe temporal dynamics could only be separated into early and late windows of 1200 ms for each motion phase (predictable or unpredictable) to compare the evolution of response time bias. However, the raw eye traces indicate the need for a higher temporal resolution of attention probe to better determine the rate of bias accumulation or de-accumulation (Figures 9 & 11). A faster rate of bias accumulation in athletes compared to the adult group may then be

identified, which the graphical results seemed to hint towards (Figure 18) even though the statistical results showed no significance.

Future directions

Covert motion tracking

The current study only examined the effect of overt motion tracking on covert spatial attention. However, objects in the visual field can also be tracked "covertly" [Morishige et al., 2021]. Whether any bias in attention occurs due to covert motion tracking still requires exploration, as well as the dynamics of such bias in response to motion predictability. Furthermore, the interaction of overt and covert motion tracking, and their combined effect on visual spatial attention should be investigated, possibly using multiple-object tracking paradigms.

Extensions into dynamic saliency models

To date, the development of dynamic saliency models of attention have focused on the extrapolation of 'low'-level image features, such as pixel color and contrast, into a motion domain using 'high'-level image features, such as motion flow estimations [Mahmood et al., 2018]. However the integration of intentional actions, such as explicit motion tracking, is a critical factor yet to be implemented as gaze scan-paths (which saliency maps attempt to predict) are task-dependant [Yarbus and Yarbus, 1967]. The model presented in this work can provide a first step to factoring in the effect of motion tracking on spatial attention for dynamic saliency, by using the bias outputs to tune the skew parameter of a skewed bi-variate Gaussian distribution (Figure 31).

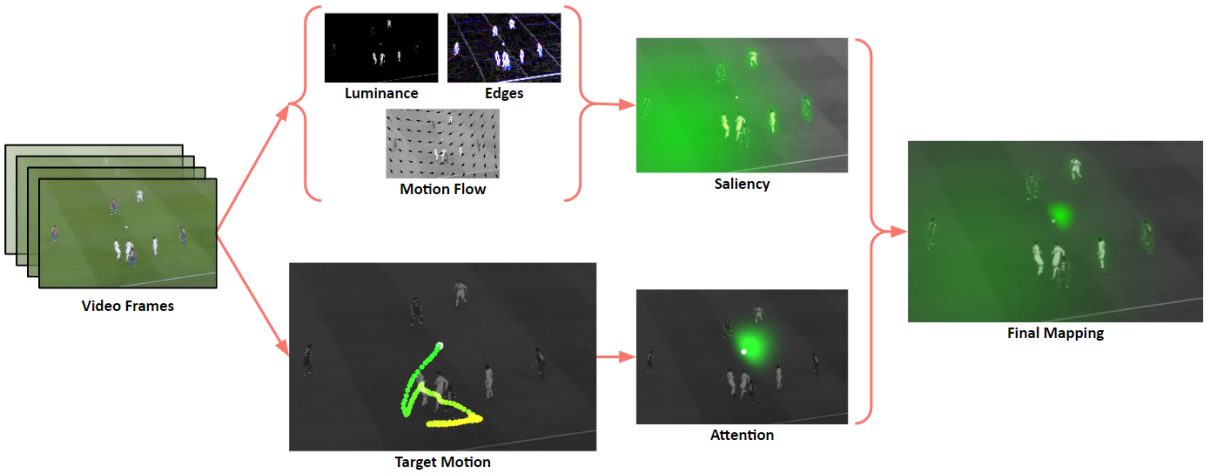

Figure 31: **Dynamic attention & saliency map:** Process of inferring attention and saliency from video frames. Image qualities such as luminance, edges, and motion flow are integrated to generate a saliency map. Target motion is extracted and attentional bias is estimated to form an attention map. Both attention and saliency are combined into a final output mapping.

The skew bi-variate Gaussian is defined as the product of a bi-variate Gaussian and the uni-variate cumulative distribution function of a spherical Gaussian:

$$f(\vec{x}, \mathbb{B}) = 2_{\phi_2}(\vec{x}, \Omega(\mathbb{B}))\Phi(\alpha(\mathbb{B})^T\vec{x}) \tag{6}$$

$$\Omega(\mathbb{B}) = \begin{bmatrix} \sigma_x = 8\mathbb{B} + 2 & 0 \\ 0 & \sigma_y = 2\mathbb{B} + 2 \end{bmatrix} \tag{7}$$

$$\alpha(\mathbb{B}) = 5\mathbb{B}^2 \tag{8}$$

Here, $2_{\phi_2}(\vec{x}, \Omega)$ is the bi-variate Gaussian, with center at $\vec{x}$ and covariance matrix of Ω defined by the estimated bias level. The diagonals of Ω scale so that horizontal and vertical variances are equal and smaller when there is no bias present, representing a tighter focus of attention about the target. Conversely, an increase in bias also increases the variance in the direction of motion which by default is left to right, but can be transformed according to the motion vector. The factor of $\Phi(\alpha(\mathbb{B})^T\vec{x})$ represents the uni-variate cumulative distribution function of a spherical Gaussian, where $\alpha((B)$ is the

skew parameter, which scales the bias value so that no bias in attention results in no skew, and larger bias values increase the skew parameter accordingly (Figure 32)).

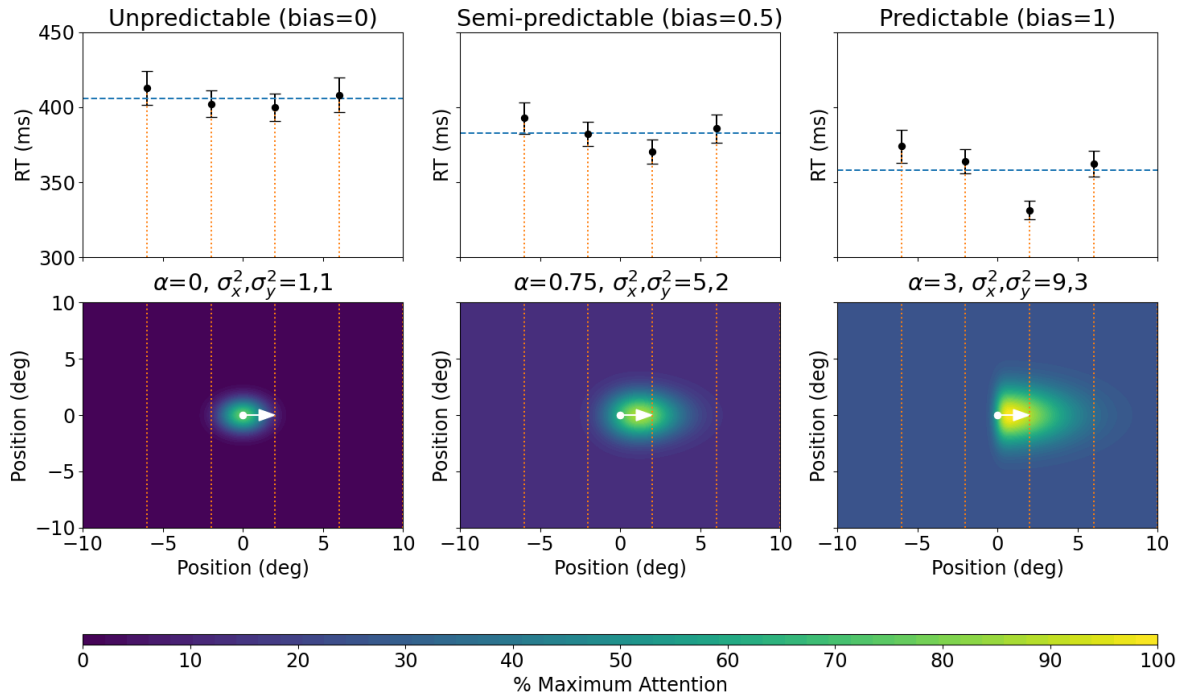

Figure 32: **Bi-variate skew Gaussian distribution generation:** The response time bias due to motion predictability (top row) and the corresponding estimation of skew (α) and variance (σ_x, σ_y) parameters to generate an attention map from a bi-variate skew Gaussian (bottom row). Target motion indicated by white arrow.

Examples of the transformation of target object motion in a dynamic attention map output, based on the computational framework presented in this work, are shown in Figures 33-36 using video inputs from team sports (e.g. ball tracking), computer games (e.g. character tracking), and e-sports (e.g. vehicle tracking).

Figure 33: **Model simulation for unbiased attention during ball tracking:** Tracking of an unpredictable ball-passing sequence; ball trajectory shown in yellow (earlier position) to green (later position) gradient; attention distribution in green is normally distributed about ball motion.

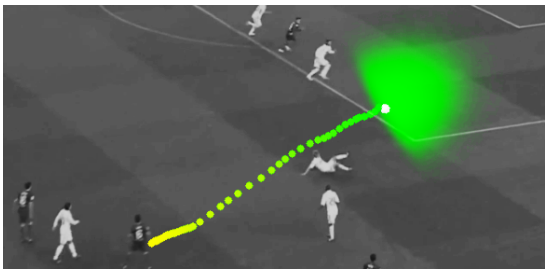

Figure 34: **Model simulation for biased attention during ball tracking:** Tracking of a predictable ball-passing sequence; ball trajectory shown in yellow (earlier position) to green (later position) gradient; attention distribution in green is skewed ahead of ball motion.

Figure 35: **Model simulation for character tracking:** Tracking of an enemy character in a computer game (*DOOM: Eternal*); character trajectory shown in yellow (earlier position) to green (later position) gradient; attention distribution in green is skewed ahead of the character's predictable jumping motion.

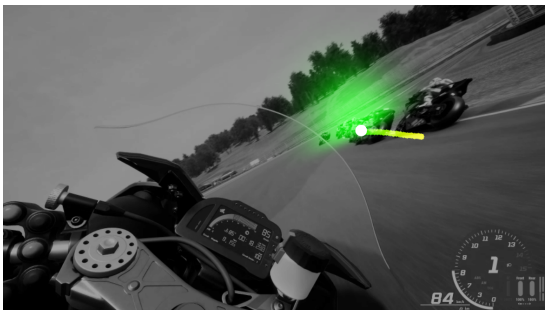

Figure 36: **Model simulation for vehicle tracking:** Tracking of competitors' motorcycles in an e-sports game (*RIDE 4*); a motorcycle is trajectory shown in yellow (earlier position) to green (later position) gradient; attention distribution in green is skewed ahead of the motorcycle's predictable motion into the turn.

Bibliography

References

[Abernethy and Russell, 1984] Abernethy, B. and Russell, D. (1984). Advance cue utilisation by skilled cricket batsmen. *Australian Journal of Science and Medicine in sport*, 16(2):2–10.

[Ando et al., 2001] Ando, S., Kida, N., and Oda, S. (2001). Central and peripheral visual reaction time of soccer players and nonathletes. *Perceptual and Motor Skills*, 92(3):786–794.

[Baars and Gage, 2018] Baars, B. and Gage, N. (2018). *Fundamentals of Cognitive Neuroscience*.

[Bailey and Lovie-Kitchin, 2013] Bailey, I. L. and Lovie-Kitchin, J. E. (2013). Visual acuity testing. from the laboratory to the clinic. *Vision research*, 90:2–9.

[Baltieri and Isomura, 2021] Baltieri, M. and Isomura, T. (2021). Kalman filters as the steady-state solution of gradient descent on variational free energy. *arXiv preprint arXiv:2111.10530*.

[Barborica and Ferrera, 2003] Barborica, A. and Ferrera, V. P. (2003). Estimating invisible target speed from neuronal activity in monkey frontal eye field. *Nature neuroscience*, 6(1):66–74.

[Barnes and Asselman, 1991] Barnes, G. R. and Asselman, P. T. (1991). The mechanism of prediction in human smooth pursuit eye movements. *The Journal of Physiology*, 439.

[Beal, 2003] Beal, M. J. (2003). Variational algorithms for approximate bayesian inference. *PhD Thesis*.

[Beauchamp et al., 2001] Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., and Haxby, J. V. (2001). A parametric fmri study of overt and covert shifts of visuospatial attention. *NeuroImage*, 14.

[Bennett et al., 2010] Bennett, S. J., Baures, R., Hecht, H., and Benguigui, N. (2010). Eye movements influence estimation of time-to-contact in prediction motion. *Experimental Brain Research*, 206.

[Berrar, 2018] Berrar, D. (2018). *Bayes' theorem and naive bayes classifier*, volume 1-3.

[Berry et al., 1999] Berry, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398.

[Bisley and Goldberg, 2003] Bisley, J. W. and Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, 299(5603):81–86.

[Blom et al., 2020] Blom, T., Feuerriegel, D., Johnson, P., Bode, S., and Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences of the United States of America*, 117.

[Brouwer et al., 2002] Brouwer, S. D., Yuksel, D., Blohm, G., Missal, M., and Lefèvre, P. (2002). What triggers catch-up saccades during visual tracking? *Journal of Neurophysiology*, 87.

[Carello and Krauzlis, 2004] Carello, C. D. and Krauzlis, R. J. (2004). Manipulating intent: Evidence for a causal role of the superior colliculus in target selection. *Neuron*, 43.

[Carrasco and McElree, 2001] Carrasco, M. and McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences*, 98(9):5363–5367.

[Cerminara et al., 2009] Cerminara, N. L., Apps, R., and Marple-Horvat, D. E. (2009). An internal model of a moving visual target in the lateral cerebellum. *The Journal of physiology*, 587(2):429–442.

[Chen et al., 2017] Chen, J., Valsecchi, M., and Gegenfurtner, K. R. (2017). Attention is allocated closely ahead of the target during smooth pursuit eye movements: Evidence from eeg frequency tagging. *Neuropsychologia*, 102.

[Clark, 2013] Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36.

[Clarke and Sokoloff, 1999] Clarke, D. and Sokoloff, L. (1999). *Circulation and Energy Metabolism of the Brain.*

[Colegatef et al., 1973] Colegatef, R. L., Hoffman, J. E., and Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics*, 14.

[Coltz et al., 1999] Coltz, J., Johnson, M., and Ebner, T. (1999). Cerebellar purkinje cell simple spike discharge encodes movement velocity in primates during visuomotor arm tracking. *Journal of Neuroscience*, 19(5):1782–1803.

[de Xivry et al., 2013] de Xivry, J.-J. O., Coppe, S., Blohm, G., and Lefevre, P. (2013). Kalman filtering naturally accounts for visually guided and predictive smooth pursuit dynamics. *Journal of neuroscience*, 33(44):17301–17313.

[Donkelaar, 1999] Donkelaar, P. V. (1999). Spatiotemporal modulation of attention during smooth pursuit eye movements. *NeuroReport*, 10.

[Donkelaar and Drew, 2002] Donkelaar, P. V. and Drew, A. S. (2002). The allocation of attention during smooth pursuit eye movements. volume 140.

[Duffy and Wurtz, 1997] Duffy, C. J. and Wurtz, R. H. (1997). Medial superior temporal area neurons respond to speed patterns in optic flow. *Journal of Neuroscience*, 17.

[Dursteler and Wurtz, 1988] Dursteler, M. R. and Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas mt and mst. *Journal of Neurophysiology*, 60.

[Eriksen and Colegate, 1970] Eriksen, C. W. and Colegate, R. L. (1970). Identification of forms at brief durations when seen in apparent motion. *Journal of Experimental Psychology*, 84.

[Eriksen and Collins, 1969] Eriksen, C. W. and Collins, J. F. (1969). Visual perceptual rate under two conditions of search. *Journal of Experimental Psychology*, 80.

[Eriksen and James, 1986] Eriksen, C. W. and James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40.

[Essen and Maunsell, 1983] Essen, D. C. V. and Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6.

[Faul et al., 2007] Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39:175–191.

[Fooken and Spering, 2019] Fooken, J. and Spering, M. (2019). Decoding go/no-go decisions from eye movements. *Journal of Vision*, 19.

[Francis and Wonham, 1976] Francis, B. A. and Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, 12(5):457–465.

[Freyd and Finke, 1984] Freyd, J. J. and Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:126–132.

[Friston, 2003] Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16.

[Fukushima, 1988] Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1.

[Gagniuc, 2017] Gagniuc, P. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.

[Gattass et al., 1981] Gattass, R., Gross, C. G., and Sandell, J. H. (1981). Visual topography of v2 in the macaque. *Journal of Comparative Neurology*, 201.

[Goodale and Milner, 1992] Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15.

[Gordon and Kravetz, 1991] Gordon, H. W. and Kravetz, S. (1991). The influence of gender, handedness, and performance level on specialized cognitive functioning. *Brain and Cognition*, 15.

[Gottlieb et al., 1994] Gottlieb, J. P., MacAvoy, M. G., and Bruce, C. J. (1994). Neural responses related to smooth-pursuit eye movements and their correspondence with electrically elicited smooth eye movements in the primate frontal eye field. *Journal of Neurophysiology*, 72.

[Hanes and Schall, 1996] Hanes, D. P. and Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274(5286):427–430.

[Harrison et al., 2012] Harrison, W. J., Mattingley, J. B., and Remington, R. W. (2012). Pre-saccadic shifts of visual attention.

[Hirano et al., 2017] Hirano, M., Hutchings, N., Simpson, T., and Dalton, K. (2017). Validity and repeatability of a novel dynamic visual acuity system. *Optometry and Vision Science*, 94.

[Hogendoorn and Burkitt, 2019] Hogendoorn, H. and Burkitt, A. N. (2019). Predictive coding with neural transmission delays: A real-time temporal alignment hypothesis. *eNeuro*, 6.

[Hu et al., 2022] Hu, D., Ison, M., and Johnston, A. (2022). Exploring the common mechanisms of motion-based visual prediction. *Frontiers in Psychology*, 13.

[Hubbard and Bharucha, 1988] Hubbard, T. L. and Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44.

[Hubel and Wiesel, 1962] Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160.

[Hubel and Wiesel, 1974] Hubel, D. H. and Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, 158.

[Hurtubise et al., 2016] Hurtubise, J., Gorbet, D., Hamandi, Y., Macpherson, A., and Sergio, L. (2016). The effect of concussion history on cognitive-motor integration in elite hockey players. *Concussion*, 1.

[Ilg, 2008] Ilg, U. J. (2008). The role of areas mt and mst in coding of visual motion underlying the execution of smooth pursuit. *Vision Research*, 48.

[Jancke et al., 2004] Jancke, D., Erlhagen, W., Schöner, G., and Dinse, H. R. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *Journal of Physiology*, 556.

[Jehee et al., 2006] Jehee, J. F., Rothkopf, C., Beck, J. M., and Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *Journal of Physiology Paris*, 100.

[Jonides, 1983] Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society*, 21.

[Kac, 1947] Kac, M. (1947). Random walk and the theory of brownian motion. *The American Mathematical Monthly*, 54.

[Kandel et al., 2000] Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science, fourth addition*, volume 4.

[Katyal et al., 2010] Katyal, S., Zughni, S., Huk, A., and Ress, D. (2010). Retinotopic maps of covert attention in human superior colliculus. *Journal of Vision*, 9.

[Keller et al., 1996] Keller, E., Gandhi, N., and Weir, P. (1996). Discharge of superior collicular neurons during saccades made to moving targets. *Journal of Neurophysiology*, 76(5):3573–3577.

[Kerzel, 2002] Kerzel, D. (2002). A matter of design: No representational momentum without predictability. *Visual Cognition*, 9.

[Khan et al., 2010] Khan, A. Z., Lefèvre, P., Heinen, S. J., and Blohm, G. (2010). The default allocation of attention is broadly ahead of smooth pursuit. *Journal of Vision*, 10.

[Khurana and Kowler, 1987] Khurana, B. and Kowler, E. (1987). Shared attentional control of smooth eye movement and perception. *Vision Research*, 27.

[Kida et al., 2005] Kida, N., Oda, S., and Matsumura, M. (2005). Intensive baseball practice improves the go/nogo reaction time, but not the simple reaction time. *Cognitive brain research*, 22(2):257–264.

[Krauzlis, 2004] Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of neurophysiology*, 91(2):591–603.

[Krauzlis et al., 2000] Krauzlis, R. J., Basso, M. A., and Wurtz, R. H. (2000). Discharge properties of neurons in the rostral superior colliculus of the monkey during smooth-pursuit eye movements. *Journal of Neurophysiology*, 84.

[Land and McLeod, 2000] Land, M. F. and McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature neuroscience*, 3(12):1340–1345.

[Lennie, 2003] Lennie, P. (2003). The cost of cortical computation. *Current Biology*, 13.

[Lovejoy et al., 2009] Lovejoy, L. P., Fowler, G. A., and Krauzlis, R. J. (2009). Spatial allocation of attention during smooth pursuit eye movements. *Vision Research*, 49.

[Mahmood et al., 2018] Mahmood, H., Islam, S. M. S., Gilani, S. O., and Ayaz, Y. (2018). Dynamic saliency model inspired by middle temporal visual area: A spatio-temporal perspective. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE.

[Martin, 2012] Martin, J. H. (2012). *Neuroanatomy Text and Atlas, 4th ed.*, volume XXXIII.

[Millidge et al., 2021] Millidge, B., Seth, A., and Buckley, C. (2021). Predictive coding: a theoretical and experimental review. *arXiv*.

[Morishige et al., 2021] Morishige, K.-i., Hiroe, N., Sato, M.-a., and Kawato, M. (2021). Common cortical areas have different neural mechanisms for covert and overt visual pursuits. *Scientific Reports*, 11(1):13933.

[Munger and Owens, 2004] Munger, M. P. and Owens, T. R. (2004). Representational momentum and the flash-lag effect. *Visual Cognition*, 11.

[Müller et al., 2003] Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., and Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *Journal of Neuroscience*, 23.

[Nachmani et al., 2020] Nachmani, O., Coutinho, J., Khan, A. Z., Lefèvre, P., and Blohm, G. (2020). Predicted position error triggers catch-up saccades during sustained smooth pursuit. *Eneuro*, 7(1).

[Nagai et al., 2002] Nagai, M., Kazai, K., and Yagi, A. (2002). Larger forward memory displacement in the direction of gravity. *Visual Cognition*, 9.

[Nassi et al., 2006] Nassi, J. J., Lyon, D. C., and Callaway, E. M. (2006). The parvocellular lgn provides a robust disynaptic input to the visual motion area mt. *Neuron*, 50.

[Nealey and Maunsell, 1994] Nealey, T. A. and Maunsell, J. H. (1994). Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex. *Journal of Neuroscience*, 14.

[Newsome et al., 1985] Newsome, W. T., Wurtz, R. H., Dursteler, M. R., and Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience*, 5.

[Nijhawan, 1994] Nijhawan, R. (1994). Motion extrapolation in catching [9]. *Nature*, 370.

[Noorani and Carpenter, 2016] Noorani, I. and Carpenter, R. (2016). The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, 64:229–251.

[Oldfield, 1971] Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, 9.

[Pack, 2001] Pack, C. C. (2001). The aperture problem for visual motion and its solution in primate cortex. *Science progress*, 84.

[Posner et al., 1980] Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109.

[Posner et al., 1984] Posner, M. I., Walker, J. A., Friedrich, F. J., and Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, 4.

[Priebe et al., 2001] Priebe, N. J., Cassanello, C. R., and Lisberger, S. G. (2001). The speed tuning of single units in macaque visual area mt depends on spatial form and contrast. *Journal of Vision*, 1.

[Rao and Ballard, 1999] Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2.

[RCoreTeam, 2021] RCoreTeam (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

[Recanzone and Wurtz, 2000] Recanzone, G. H. and Wurtz, R. H. (2000). Effects of attention on mt and mst neuronal activity during pursuit initiation. *Journal of Neurophysiology*, 83.

[Reed and Vinson, 1996] Reed, C. L. and Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 22.

[Riesenhuber and Poggio, 1999] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2.

[Rottach et al., 1996] Rottach, K. G., Zivotofsky, A. Z., Das, V. E., Averbuch-Heller, L., Discenna, A. O., Poonyathalang, A., and Leigh, J. R. (1996). Comparison of horizontal, vertical and diagonal smooth pursuit eye movements in normal human subjects. *Vision Research*, 36.

[Schall, 2004] Schall, J. D. (2004). On the role of frontal eye field in guiding attention and saccades. *Vision research*, 44(12):1453–1467.

[Schall et al., 1995] Schall, J. D., Hanes, D. P., Thompson, K. G., and King, D. J. (1995). Saccade target selection in frontal eye field of macaque. i. visual and premovement activation. *Journal of Neuroscience*, 15.

[Schoenfeld et al., 2002] Schoenfeld, M., Heinze, H.-J., and Woldorff, M. (2002). Unmasking motion-processing activity in human brain area v5/mt+ mediated by pathways that bypass primary visual cortex. *NeuroImage*, 17.

[Schreiber et al., 2006] Schreiber, C., Missal, M., and Lefèvre, P. (2006). Asynchrony between position and motion signals in the saccadic system. *Journal of Neurophysiology*, 95(2):960–969.

[Seya and Mori, 2012] Seya, Y. and Mori, S. (2012). Spatial attention and reaction times during smooth pursuit eye movement. *Attention, Perception, and Psychophysics*, 74.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27.

[Shi et al., 1998] Shi, D., Friedman, H. R., and Bruce, C. J. (1998). Deficits in smooth-pursuit eye movements after muscimol inactivation within the primate's frontal eye field. *Journal of Neurophysiology*, 80.

[Sillito et al., 1994] Sillito, A. M., Jones, H. E., Gerstein, G. L., and West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369.

[Sokhn et al., 2019] Sokhn, N., Wuilleret, A., and Caldara, R. (2019). Go/no-go saccadic reaction times towards visual field targets differ between athletes and nonathletes. In *2019 11th International Conference on Knowledge and Smart Technology (KST)*, pages 222–225. IEEE.

[Solari et al., 2015] Solari, F., Chessa, M., Medathati, N. V., and Kornprobst, P. (2015). What can we expect from a v1-mt feedforward architecture for optical flow estimation? *Signal Processing: Image Communication*, 39.

[Spering et al., 2011] Spering, M., Schütz, A. C., Braun, D. I., and Gegenfurtner, K. R. (2011). Keep your eyes on the ball: Smooth pursuit eye movements enhance prediction of visual motion. *Journal of Neurophysiology*, 105.

[Spratling, 2012] Spratling, M. W. (2012). Predictive coding accounts for v1 response properties recorded using reverse correlation. *Biological Cybernetics*, 106.

[Stark and Bridgeman, 1983] Stark, L. and Bridgeman, B. (1983). Role of corollary discharge in space constancy. *Perception & Psychophysics*, 34.

[Steinmetz and Constantinidis, 1995] Steinmetz, M. and Constantinidis, C. (1995). Neurophysiological evidence for a role of posterior parietal cortex in redirecting visual attention. *Cerebral Cortex*, 5(5):448–456.

[Tanaka and Lisberger, 2001] Tanaka, M. and Lisberger, S. G. (2001). Regulation of the gain of visually guided smooth-pursuit eye movements by frontal cortex. *Nature*, 409.

[Tanaka and Lisberger, 2002] Tanaka, M. and Lisberger, S. G. (2002). Role of arcuate frontal cortex of monkeys in smooth pursuit eye movements. i. basic response properties to retinal image motion and position. *Journal of Neurophysiology*, 87.

[TP et al.,] TP, G., Mehta, H., Gokhale, P., and Shah, C. A comparative study of visual reaction time in basketball players and healthy controls.

[van Heusden et al., 2019] van Heusden, E., Harris, A. M., Garrido, M. I., and Hogendoorn, H. (2019). Predictive coding of visual motion in both monocular and binocular human visual processing. *Journal of Vision*, 19.

[Verfaillie and d'Ydewalle, 1991] Verfaillie, K. and d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17.

[Xivry and Lefèvre, 2007] Xivry, J. J. O. D. and Lefèvre, P. (2007). Saccades and pursuit: Two outcomes of a single sensorimotor process. *Journal of Physiology*, 584.

[Yang et al., 2012] Yang, J., Lee, J., and Lisberger, S. G. (2012). The interaction of bayesian priors and sensory data and its neural circuit implementation in visually guided movement. *Journal of Neuroscience*, 32(49):17632–17645.

[Yantis and Jonides, 1984] Yantis, S. and Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10.

[Yarbus and Yarbus, 1967] Yarbus, A. L. and Yarbus, A. L. (1967). Eye movements during perception of complex objects. *Eye movements and vision*, pages 171–211.

[Yee, 2017] Yee, A. (2017). Investigation of vision strategies used in a dynamic visual acuity task.

[Zeki, 1980] Zeki, S. (1980). The response properties of cells in the middle temporal area (area mt) of owl monkey visual cortex. *Proceedings of the Royal Society of London - Biological Sciences*, 207.

[Zeki, 1974] Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology*, 236.