COMPARING STATISTICAL METHODS FOR INFERRING CONTRIBUTIONS OF VISUAL ONLINE CONTROL FROM HUMAN LIMB TRAJECTORIES

by

Ghislain d'Entremont

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
October 2018

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Visual motor control involves using visual information about the limb and the target to adjust the trajectory of the limb towards the target to improve movement accuracy. The primary objective of the thesis was to demonstrate that improvements to the standard methods of statistical analysis of movement trajectory data can substantially improve the quality of the inferences made about those data. A Bayesian hierarchical gaussian process regression (GPR) model was compared to traditional analysis techniques in its ability to accurately estimate experimental effects. Analyses were run on experimental data collected from a basic vision/no-vision goal-directed reaching task, and simulated data from theoretically plausible generative model. Broadly, the expected experimental effects of vision were generated. The Bayesian hierarchical GPR method was successfully implemented and conferred some substantial benefits in contrast to many of the traditional methods. However, given several usability limitations, the Bayesian hierarchical GPR method may be best used as a specialty tool for statistically savvy researchers seeking to maximize the inferential capacity of their analysis of movement trajectories.

LIST OF ABBREVIATIONS USED

GPR	Gaussian Process Regression
DV	Dependent Variable
RT	Reaction Time
MT	Movement Time
CE	Constant Error
VE	Variable Error
TAPV	Time After Peak Velocity
TPV	Time to Peak Velocity
TPD	Time to Peak Deceleration
IECE	Index of Error Correction Effectiveness
rANOVA	Repeated Measures ANOVA
FFT	Fast Fourier Transform
fANOVA	Functional ANOVA
RM	Repeated Measures
GP	Gaussian Process
NHST	Null Hypothesis Significance Testing
IRED	Infra-Red Emitting Diodes
SD	Standard Deviation
η_G^2	Generalized Eta Squared
FLSD	Fisher Least Significant Difference
PA	Peak Acceleration
PV	Peak Velocity
PD	Peak Deceleration
END	End of Movement
MCMC	Markov Chain Monte Carlo

CHAPTER 1 INTRODUCTION

1.1 VISUAL CONTROL OF MOVEMENT

One focus in the field of motor control is the study of sensorimotor and cognitive processes that contribute to goal-directed upper limb movements. In the sensorimotor domain, of interest is how the actor uses different types of sensory feedback (e.g. vision, proprioception) during the movement to improve movement accuracy. Online control refers to processes that occur during the movement. Primarily, online control encompasses the use of sensory and proprioceptive feedback regarding the limb, the target, and the environment to adapt a given movement trajectory. Conversely, offline control refers to the processes that occur before and after the movement. Forms of offline control include generating a model to predict how the movement will unfold, incorporating terminal feedback to produce future movements (learning), and adapting motor control strategies based on knowledge about the task.

Most models of online control have posited two control phases: an early phase during which the majority of the distance to the target is covered (i.e. during the primary movement), and a later phase during which movement errors are corrected based on sensory information (for history and discussion; see Elliott, Helsen, &, Chua, 2001). Elliott et al. (2017) define these two online control phases as *impulse control* and *limb-target control*.

Impulse control involves the early and continuous comparison of expected and perceived sensory consequences (e.g. visual and proprioceptive information). In other words, impulse control consists of the rapid trajectory adjustments made on the basis of discrepancies between the expectancies of an internal representation of the limb trajectory and the actual feedback regarding that trajectory. This corrective process occurs throughout the primary (distance-covering) movement that generally encompasses the peak acceleration and deceleration kinematic markers. Alternatively, the limb-target phase of the movement involves the comparison of the relative spatial positions of the target and limb, later in the movement. This form of online control entails trajectory adjustments in cases where impulse control was insufficient in minimizing the error between the primary movement and target endpoints.

Current models of goal-directed reaching such as the one proposed by Elliott et al. (2017) rely on a large literature of motor control research. Experiments focused on *visual* motor control, in particular, tend to be structured in the following way. First, participants rest their index finger at the bottom of a touch screen. Upon cue onset, participants reach towards a target, which will have appeared on the touch screen, as quickly and as accurately as possible. Common experimental manipulations in these sorts of studies involve (1) removing vision of limb and/or target at any point during a trial, (2) changing the location of the target or perturbing the limb at any point during the movement, (3) manipulating the information participants have about the nature of a trial prior to movement onset (i.e. changing the feedback schedule). In the following few paragraphs we will summarize foundational studies in visual motor control to give the reader concrete examples of the way in which experiments are designed to answer specific

questions about visual motor control.

Real-time visual information about the movement environment is used to adapt and correct the movement when the limb is in flight. A half-century ago (1968), Keele & Posner concluded that it takes 190-260 ms to for a human to process visual feedback. Their design involved motivating participants to make accurate reaching movements to a small target using different movement times. On half the trials, the lighting in the room was eliminated upon movement onset. The authors compared the number of target misses between the vision and no-vision conditions across all movement time conditions. They posited that the movement time condition for which there began to be an accuracy benefit of vision compared to no-vision corresponded to the time required for humans to process visual feedback.

Knowledge about the availability of visual feedback affects the relative contributions of online and offline control processes (Khan, Chua, Elliott, Coull, & Lyons, 2002). Unlike Keele & Posner (1968), in which all vision and no-vision trials were randomly ordered, Khan et al. (2002) also included some blocked conditions in which all trials within a given block were *either* vision *or* no-vision trials. When participants knew they were going to receive visual feedback, they spent less time planning the movement and more time adjusting the movement after peak limb deceleration, compared to the random feedback condition where they did not know whether they would receive visual feedback. In other words, participants likely prioritized online control, in particular limb-target control, over offline control processes when they had knowledge that they would receive visual feedback. These results show that motor control strategies differ depending on the

nature of the information given about the task.

The visual information near the target is integrated into the movement plan both prior to and during the movement. Whitney, Westwood, & Goodale (2003) demonstrated that rapid movement trajectories to stationary targets are affected by visual motion in close proximity to the target. The authors determined that trajectories to targets were biased in the direction of the motion of a nearby task-irrelevant stimulus (i.e. vertically shifting patterns). Whitney et al. (2003) concluded that the visuomotor system makes use of sensory motion signals, continuously, in both offline and online stages of motor control.

Only a small subset of visual motor control studies has been summarized above. Of course, more work has been and continues to be done in this field to determine how humans use sensory information to adjust their movements online to produce accurate movements. The focus of the thesis, however, was not to directly contribute to this body of work, but instead to examine the statistical tools that are currently used to make sense of the rich data sets that are commonly gathered in human limb trajectory research. Improved statistical techniques will help researchers build on the history of using movement trajectories as a metric to examine the use of sensory feedback and the influences of cognitive processes on human movement.

The primary objective of the thesis was to demonstrate that improvements to the methods of statistical analysis of trajectory data can confer substantial benefits with respect to a researcher's ability to detect real experimental effects and maximize the nuance with which those effects can be described. This objective was pursued in the

context of visual motor control research because many competing methods of analysis have been proposed in this field, indicating the important role statistical inference plays in resolving scientific hypotheses in the domain of visual motor control

1.2 CRITICAL APPRAISAL OF COMMON ANALYSIS METHODS

Although experimental design choices play a large part in determining which physiological process of motor control is under investigation, statistical analysis decisions also play a substantial role. Human limb movement experiments naturally provide rich data sets that include 3 dimensional trajectories through time across multiple trials, participants, and conditions. Therefore, the choice of analysis methods can have a large impact on the interpretation of the results.

Figure 1 contains sample displacement, velocity, and acceleration profiles for movement amplitude (i.e. away from the participant). Commonly, the displacement profiles are measured directly by tracking the position of the participant's limb through time with an optoelectronic device. The velocity and acceleration profiles, however, tend to be calculated by differentiating (once or twice) the displacement profile. Differentiation can amplify the measurement noise present in the displacement profile, but this issue can largely be mitigated using proper filtering procedures (for discussion, see Hansen, Elliott, & Khan, 2007).

In the figure, the two control phases (as described in subsection 1.1) are labelled as initial impulse and error correction instead of impulse control and limb-target control, although

they largely represent the same underlying control mechanisms described by Elliott et al. (2017). It can be seen that the initial phase constitutes a large portion of the displacement profile. The second, error-correcting, phase is made up of discrete changes in displacement at the end of the movement and is defined as the time period following zero-line crossings in the velocity (positive to negative) or acceleration profile (negative to positive), or acute changes in the acceleration profile. Figure 1 showcases kinematic profiles corresponding to a single dimension of space (i.e. movement amplitude). However, limb position is often recorded in 3D. Therefore, researchers often run analyses on the properties of these kinematic profiles in select dimensions, independently. Otherwise, they tend to select properties of resultant kinematic profiles. The specifics of how and what analyses are commonly run on kinematic profiles to answer scientific hypotheses will largely be the focus of this sub-section.

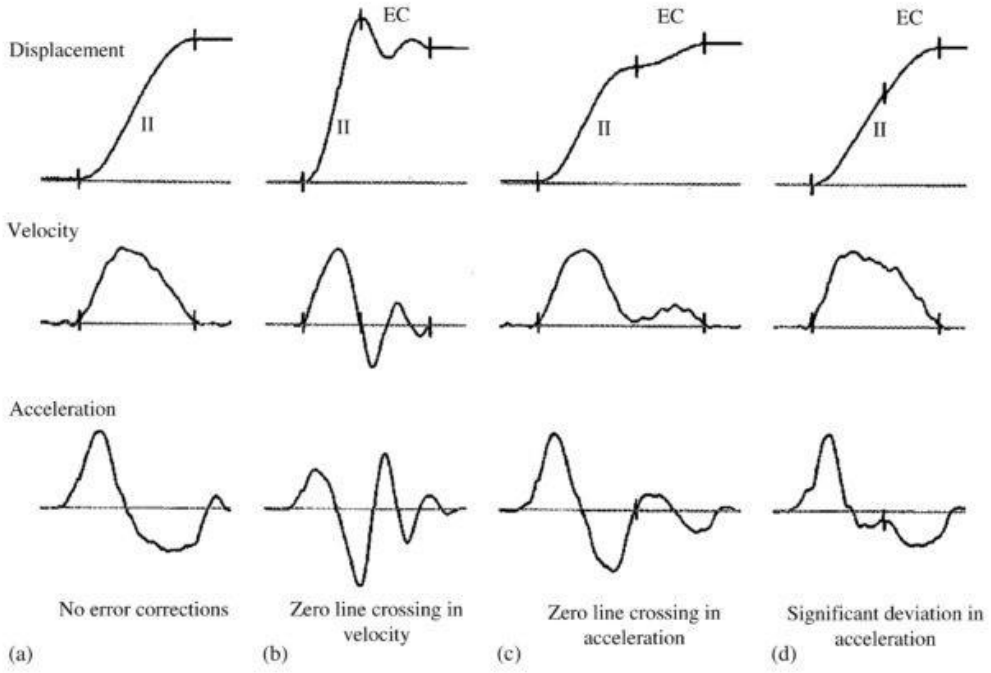


Figure 1 Sample displacement, velocity, and acceleration profiles for movement amplitude (i.e. the same direction as the limb - typically away from the body). The segments, demarcated by vertical lines through the displacement profiles and labelled as either II or EC, represent the *initial impulse* and *error correction* phases of the movement, respectively. These phases map onto the impulse control and limb-target control phases posited by Elliott et al. (2017), respectively. *Adapted from Khan et al. (2006) which was adapted from Khan et al. (1998).*

Researchers have been hard-pressed to find consensus on the topic of which analyses are most appropriate for given research questions. Perhaps one of the most common analyses is the use of repeated measures ANOVAs (rANOVAs) on several outcome measures as dependent variables (DVs): reaction time (RT), movement time (MT), and movement error (for examples, see Heath, Westwood, & Binsted, 2004; Heath, 2005; and Khan et al., 2002). RTs are simply the time interval between the onset of the cue indicating the that the participant can begin the movement and the onset of the movement itself. MTs

are the time interval between the onset of the movement and the offset of the movement (or, when the participant reaches the target on screen). Movement error can be quantified in several ways. For instance, one could take the distance between the index finger and the center or periphery of the target at the time that the participant makes contact with the touch screen. This is commonly referred to as constant error (CE). Alternatively, one can compute variable error (VE) which typically involves taking the standard deviation of the distance between the target and the index finger of the participant over several trials.

Another popular set of analyses is that of kinematic measures to infer information regarding the use of sensory feedback. For instance, authors will often apply rANOVAs to the following DVs to evaluate trajectory symmetry across conditions: time after peak velocity (TAPV), time to peak velocity (TPV), and time to peak deceleration (TPD; for examples, see Heath et al., 2004; and Heath, 2005). The computation of these DVs is fairly self-evident. For instance, TAPV would simply be the time interval between the sample at which the participant reached peak limb movement velocity and the sample at which either the movement terminated or the participant reached the touch screen. An increase in TAPV is typically interpreted as an indication that the participant is spending more time making online corrections.

Another category of measures attempts to quantify the discrete corrections within movement conditions. For instance, researchers may quantify trajectory discontinuities by counting the number of substantive changes in the velocity or acceleration profiles of the trajectories (for discussion, see Elliott & Hansen, 2010). In other instances, researchers

have created online motor control indices such as the index of error correction effectiveness (IECE). The IECE is meant to indicate the degree to which a discrete error correction submovement was effective in decreasing the actual error of the entire movement (for discussion, see Elliott & Hansen, 2010).

Generally, for these types of kinematic measurements, multiple DVs are analyzed. Each DV is submitted to a rANOVA, and each rANOVA is followed-up by post-hoc tests. The problem with this sort of analysis routine, particularly when the number of DVs (and therefore statistical tests) is large, is that of multiple comparisons. Specifically, uncorrected multiple comparisons inflate type I error rates, whereas corrected multiple comparisons inflate type II error rates. Further, running separate rANOVAs on DVs that are potentially related may result in an incomplete understanding of the results. For instance, if one fails to assess the relationship among DVs, then one may fail to see the way in which their effects are manifestations of an underlying cognitive or perceptual mechanism, and may therefore describe the pattern of results in a more fragmented manner. In summary, the relationship among DVs should be explicitly modelled and/or DVs should be collapsed into parsimonious measures of visual motor control regulation.

Spatial variability profiles are another common method of analysis of movement trajectories (for examples, see Khan et al., 2002; and Khan et al., 2006). Spatial variability profiles are calculated by taking the standard deviation of the position of the limb across trials at several points in time. Typically, these time points are kinematic markers (e.g. peak acceleration, peak velocity, peak deceleration, etc.). The computed

standard deviations are then submitted to rANOVAs as levels of a time factor, sometimes broken down further by experimental conditions. The premise of this method of analysis is that online corrections will cause trajectories to deviate from the planned movement and therefore result larger amounts of variability in the trajectories across trials (i.e. the planned movement trajectory is more stable, less variable, on a trial-by-trial basis).

While spatial variability profiles reduce the issue of performing multiple rANOVAs by including timepoints or kinematic marker as a factor in the analysis, a different problem arises. Standard deviations are, by definition, bounded by zero and therefore may not satisfy the distributional assumptions of rANOVAs. Since the sample size in human limb trajectory studies often does not exceed 20, the central limit theorem does not obviously apply (i.e. the sampling distribution of standard deviations in this context is not necessarily approximately normal).

An even larger problem is that spatial variability profiles fail to make use of the entire trajectory and do not account for the *temporal* variability associated with the kinematic markers themselves, as pointed out by Hansen, Elliott, & Khan (2008). The authors addressed at least the former shortcoming by developing a 3D variability profile method that, after normalizing with respect to time, made use of most of the trajectory information. Their method involved specifying an ellipsoid for each normalized time point, with the standard deviation of the position of the limb across trials, in each of three coordinates, mapping onto the radii of the ellipsoid. The volume of each ellipsoid is calculated to obtain a spatial variability profile (i.e. ellipsoid volume as a function of

normalized time). It is unclear whether Hansen, Elliott, & Khan (2008) truly solve the problem of temporally unstable kinematic markers by getting rid of them all together. To the extent that these markers are special in their relation to physiological processes, it may at least be useful to indicate where they occur along the normalized time course of the 3D spatial variability profiles. More importantly, this novel method does not account for the autocorrelation within the spatial variability waveform. Finally, the method proposed by Hansen, Elliott, & Khan (2008) does not solve the problem presented above regarding the potential violation of the rANOVAs distributional assumptions.

R^2 analyses have also been espoused in realm of visual motor control research (for examples, see Heath et al., 2004; and Heath, 2005). These involve running a standard linear regression on the relationship between the position of the limb at the end of the movement and its position at a certain point during the movement, for each participant, and then running rANOVA analyses on the R^2 values from those linear regression analyses. The R^2 method is a rather crude method for estimating the degree to which online control affects the relationship between the hand's position early in movement and its position later in the movement. The richness of the information garnered from the R^2 method could be greatly improved with modern methods. If the goal is to evaluate the degree of autocorrelation in the trajectory, then the level of autocorrelation should be explicitly estimated. Alternatively, if the goal is to account for the number of discontinuities attributable to online control, then those discontinuities should be modelled explicitly.

Out of the more standard methods summarized in the previous section, the variability

profile analysis appears to be the best measure of online control (Elliott & Hansen, 2010; and Khan et al., 2006). Recently, novel applications of statistical methods have been presented. For instance, Grosbois & Tremblay (2016) applied rANOVAs on the fast fourier transforms (FFTs) of the residuals of polynomial fits of trajectories, thereby conducting an analysis in the frequency domain. This approach introduces *experimenter degrees of freedom* (the set of decisions regarding data collection and analysis that researchers possess) in (1) choosing the polynomial order, which may not always be particularly obvious in each application, and (2) choosing the frequency bins for statistical analyses. Though these decisions could be guided by the research interest of the experimenter and prior knowledge, each decision simultaneously and unnecessarily reduces model flexibility and parameter estimate uncertainty. Furthermore, reducing a power spectrum to bins for the purpose of submitting the data to a rANOVA results in a further loss of information. Ideally, the entire power spectrum of the model residuals could be modelled directly.

Another novel method is that of Gallivan & Chapman (2014). They applied repeated measures (RM) functional ANOVAs (fANOVAs) to spline fitted trajectories. Although this method was not presented in the context of visual control paradigms, it *does* generalize to any endeavor to assess differences among trajectories belonging to different conditions. Unfortunately, splines lack some flexibility as an arbitrary number of knots have to be specified (Gelman et al., 2014). In particular, the issue with this choice of model hyperparameters (number of knots) is that it removes the uncertainty associated with choosing those very hyperparameters. Also, improperly specifying those hyperparameters could result in inaccuracies in the model.

Having laid out specific criticisms for most of the standard methods of analysis, some limitations applicable to the bulk of these methods will be stated. Generally, the hierarchical structure of the data is not considered. One should model the mean trajectory information from a given participant as having been sampled from some population distribution, and then model trial-wise trajectory information regarding that participant as having been sampled from that participant's unique distribution. A misspecified model can result in the failure to adjust estimates based on having multiple trials per participant, the failure to account to imbalances in number of trials across participants, and the unjustified removal of variation at the participant level (for discussion, see McElreath, 2015). Furthermore, hierarchical models facilitate the examination of the relationship among parameters. For instance, there may be a correlation between the main effect of having vision during a movement and the main effect of not having vision during a movement. Correlation structures such as these can easily be incorporated into hierarchical models.

In summary, many commonplace analysis procedures suffer from some combination of the following limitations: they analyze several DVs per experiment without accounting for the relationship among them; they run a large number of statistical tests, thereby unnecessarily inflating type I and/or type II error rates; they do not use statistical models for which the distributional assumptions for a given DV are necessarily met; they do not make use of the full trajectory and thus may lose valuable information; and they do not account for the natural hierarchical structure of the data. The purpose of this thesis was to

propose a method that is largely an antidote to these limitations. A large part of that statistical solution will be described in the following subsection.

1.3 GAUSSIAN PROCESSES

Gaussian Processes (GPs) will offer a partial solution to some of the common analysis limitations described above. Before describing the way in which GPs will improve statistical inference, a brief GP tutorial will be given in theory and in practice. *Processes* are the generalization of probability distributions to functions. Making them *Gaussian* facilitates computations. Specifically, a GP is a collection of random variables for which any subset can be described by a multivariate gaussian distribution (for foundational introduction to GPs, see Rasmussen & Williams, 2006). GPs can be used to model time series data of arbitrary form and shape (unlike polynomial regression techniques which require prespecified structure). GPs are almost entirely defined by their covariance function (i.e. autocorrelation structure). Commonly, the covariance function is determined by two parameters: the volatility parameter controls the 'wiggliness' (or, conversely, smoothness) of the function, and the amplitude parameter controls the scale of the function. These are often referred to as *hyperparameters*. Figure 2 illustrates the effect of these hyperparameters on the shape of the functions that can be sampled from a GP with a covariance function defined by these hyperparameters. For now, it is sufficient to understand that GPs can be used to generate estimates (with uncertainty) of the trajectory itself, and of any of its derivatives. Estimates of the kernel hyperparameters (i.e. volatility and amplitude) can also be modelled.

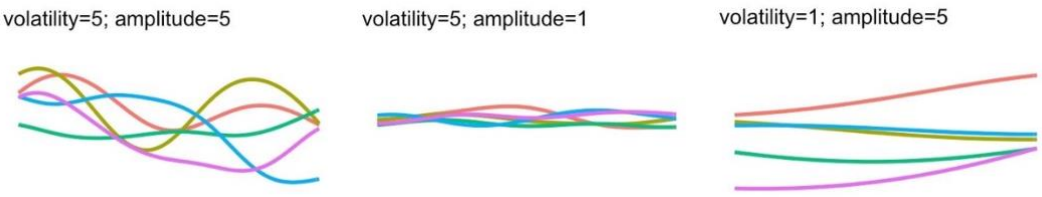


Figure 2 Random samples from a GP with an exponentiated quadratic kernel with volatility hyperparameters of 5 (left & middle) and 1 (right), and amplitude hyperparameters of 5 (left & right) and 1 (middle). The mean function is set at zero (i.e. the average of the functions drawn from this GP will tend towards zero). Each sampled function is made up of 101 points, which go 1 to 101 with equal spacing (these details matter for the absolute value of the volatility and amplitude hyperparameters to be meaningful). The y-axis is on the same scale in all three graphs.

Many of the analysis critiques described in previous subsection can simultaneously be addressed by GPs. Generally, GPs allow for the comprehensive analysis of the entire trajectory. Comparisons can be made among the population mean trajectories of different conditions and their derivatives. Further, one can interpret differences among covariance hyperparameters stemming from different conditions. To the extent that the visual control information contained in some of the numerous standard DVs are overlapping and/or contained in the full trajectories or their derivatives, GPs would provide a more holistic, intuitive, and rich method of analysis. For example, if differences in visual control between two conditions are manifested in the difference between the displacement, velocity, or acceleration profiles of the two conditions, then GP estimates of these profiles will indicate when and how many of these profiles differ between conditions. If the difference between two experimental conditions is expected to manifest itself more abstractly (e.g. one trajectory is expected to be 'wigglier' than another) then estimates (with uncertainty) of the hyperparameters of the trajectories, and their derivatives, can be

compared.

The variability profiles of the trajectories can be modelled continuously using GPs. The estimates are also more sensible in principle as they represent the entire variability profile as a function, whereas standard spatial variability profile and epsilloid analyses breakdown the time into a multi-level factor thereby failing to account for the relationship between spatial variability values in temporal proximity. In fact, using the modelling software and methods to be described later on, multiple GPs (e.g. trajectories, their derivatives, and variability profiles) can be modelled simultaneously. In terms of assessing autocorrelation examining covariance function parameters and/or specific covariance values would provide information regarding how early movement patterns predict later ones. For instance, differences in volatility hyperparameters would indicate that early parts of a movement predict later parts of the movement more so in one condition over another. One could also determine autocorrelation between two specific points in the movement by examining the covariance matrix derived from the covariance function and its hyperparameters. These measures of autocorrelation could overlap with results from the R^2 analysis. Finally, as will be described in the methods section, GPs can easily be incorporated into a hierarchical framework.

GPs can also examine frequencies, similarly to the analysis performed Grosbois & Tremblay (2016), by decomposing trajectories into sub-components oscillating at certain frequencies (for example of such an application, see Gelman et al., 2014). However, that is beyond the scope of the current thesis.

The other novel method involved using splines to model full limb trajectories (Gallivan & Chapman, 2014). The benefit of GPs over splines is that they do not require the specification of hyperparameters (number of knots) ahead of time (Gelman et al., 2014). Beyond the modelling of the functions themselves, there is also the question of how statistical inferences will be made on these functions. In contrast to Gallivan & Chapman (2014), who suggested the use of functional data analysis on spline fitted models, fairly recent software developments allow for the incorporation of the model fitting procedure along with the inference procedure in a flexible manner. Largely, this will be the subject of the following section.

1.4 BAYESIAN INFERENCE AND GAUSSIAN PROCESS REGRESSION

In practice, GPs are specified as *prior distributions* for arbitrary regression functions (for concise conceptual summary and application example, see Gelman et al., 2014). The process of inferring the form of the regression function, having assumed that function to be a GP, is referred to as Gaussian Process Regression (GPR; Cox, Kachergis, & Shiffrin, 2012). GPR is inherently a Bayesian style of inference as it involves combining information about the prior and the data (with respect to the likelihood) to generate a posterior distribution.

Estimating model parameters in fully Bayesian models generally cannot be done analytically. Instead, computationally intensive approximations need to be used. Since the

early applications of GPR, computational limitations have restricted the fully Bayesian estimation of the covariance function parameters (Rasmussen & Williams, 2006), the consequences of which include the incomplete characterization of the uncertainty in the model and inability to make inference at the level of the hyperparameters (Flaxman, Gelman, Neill, & Wilson, n.d.). Thankfully, recent advancements in approximation methods and software enables researchers to implement fully Bayesian GPR (Flaxman et al., n.d.; Stan Development Team, 2017).

Bayesian inference is not just useful for the implementation of GPR, it is arguably a preferable method of analysis compared to frequentist methods. Kruschke & Liddell (2016) make the case that "the New Statistics" that emphasize parameter estimation instead of null-hypothesis significance testing are better served by Bayesian methods than they are by frequentist methods. Bayesian parameter estimation provides the richest source of information regarding the parameters of interest in the form of posterior distributions, which are explicit, continuous representations of the relative credibility of parameter values (for a comprehensive introduction to Bayesian parameter estimation, see Kruschke, 2014).

In contrast, frequentist methods offer parameter estimates in the form of confidence intervals. A confidence interval is the range of values that one must believe contains the parameter of interest if one is to be correct about such a statement a given percentage of the time (typically, 95% of the time). Unfortunately, these long-run properties are not intuitive. Indeed, many researchers struggle to interpret confidence intervals properly (for

discussion, see Cumming, & Finch, 2005). Bayesian credible intervals (some interval of the posterior distribution) have the property of containing the true parameter value with a given probability. Therefore, one could say about a 95% credible interval that one believes that there is a 95% probability that the true parameter value is within it. As noted above, this intuitive interpretation would not be a proper interpretation of a confidence interval.

Confidence intervals also lack distributional information. For example, parameter values closer to the center of a confidence interval are not necessarily more probable than values at the extremities of the interval. Bayesian posterior distributions, which tend to have an approximate normal distribution when not bounded, are continuous representations of the relative credibility of parameter values. In other words, if the density at one parameter value is greater than that of another parameter value then the former is believed to be more probable than the latter. The result of this is that values that tend to be closer to the center of a posterior distribution or a credible interval also tend to be more credible parameter values.

The reader may have noticed that the terms ‘credibility’ and ‘probability’ are being used almost interchangeably. The distinction is a subtle one. Since the prior distribution in Bayesian inference is rarely completely empirical (i.e. since the experimenter’s beliefs about the parameter values have some influence on the prior distribution selected), it is not technically correct to refer to the posterior distribution, which is derived from the prior in part, as a probability distribution – although, it does have its properties. Thus, a

credible interval is the interval that represents one's belief about the probability with which the true parameter value lies within the interval. Bayesian posterior distributions will henceforth be described with probabilities instead of credibilities to avoid confusion even though they are technically representations of relative credibility.

Bayesian inference is a natural antidote to a lot of the problems and limitations that arise with frequentist approaches. For instance, null hypothesis significance testing (NHST) methods require that the intended number of participants be sampled for the p-values to have their desired properties, whereas the Bayesian posterior is not affected by such sampling intentions as long as those intentions are not conditional on the statistical results (Dienes, 2011; Kruschke, 2012). However, even in the case where one is explicitly fishing for positive results, Bayesian and frequentist methods behave rather differently. By simulating the most egregious cases of statistical wrongdoing, sequential testing leads to 100% type I error rates in the frequentist context. In other words, if one runs a statistical test after each incremental increase in sample size until a significant finding arises, then one will always attain significant finding given enough time (for demonstration, see Kruschke, 2012). In the very same simulation, Kruschke (2012) observed sub-10% type I error rates using Bayesian inference.

In the frequentist context, researcher intentions, subsequent to the completion of data collection, commonly compromise statistical inference (Dienes, 2011; Kruschke, 2014). In particular, multiple comparisons and post-hoc (vs. planned) tests tend to inflate type I error rates. These complications ultimately stem from the fact that researcher intentions,

in affecting the scope of things that could have occurred, modify the long-run frequencies of test statistics. Imagine that a researcher runs a pairwise test to examine the difference between the two groups, among a plethora of groups, that appear to manifest the largest effect. The fact that the researcher may not have run that statistical test, had the relevant pairwise comparison not constituted the largest mathematical difference between two groups among the possible pairwise comparisons, affects the true properties of the test statistics involved in the statistical comparisons. In this way, the intentions of the researcher must be considered in evaluating all the results that could have materialized (i.e. all the tests that could have been run) had a different sample been drawn. Indeed, this is why researchers are often encouraged to plan their statistical pipeline ahead of time. The practical negative consequences of researchers having access to these experimenter degrees of freedom have been substantiated (Simmons, Nelson, & Simonsohn, 2011[1]) and are often distilled to type I error rate inflation. The Bayesian posterior is not dependent on the researcher's testing intentions in these ways.

[1] Though these authors explicitly state that they believe that Bayesian inference is not a solution to the problem they highlight, their objections - Bayesian methods themselves and the choice of prior distributions contribute to the experimenter degrees of freedom - are cursory and weak. For one, the decision to adopt frequentist or Bayesian analysis techniques should occur prior to the presentation of the data, on a matter of principle, and is likely to be a lasting decision. Indeed, this is the case herein. Secondly, the establishment of norms regarding the specification of prior distributions has occurred (Gelman et al., 2014; Stan Development Team, 2017) and is a worthwhile endeavor given the positive benefits described in this thesis.

A narrative example in the context of visual control research of how researcher intentions can have differential effects depending on the statistical theory used (frequentist vs. Bayesian) follows over the next four paragraphs. Imagine a case where a researcher collects limb trajectory data for two experimental conditions using a within-subjects design. The experimental conditions are randomly intermixed across four blocks. The researcher is interested whether participant trajectories are more variable at a given kinematic marker in one condition vs. another. Therefore, the researcher creates a spatial variability profile which evaluates the variability at four kinematic markers and compares these profiles between conditions. Further, the researcher segments the data into blocks to examine the effect of training. Thus, a 4 (block) x 2 (condition) x 4 (kinematic marker) rANOVA is required. The researcher is also interested in whether outcome measures such as RT, MT, and end-point error differ between conditions and/or among blocks. Therefore, a 2 (condition) x 4 (block) rANOVA is required for each of the three DVs.

The researcher runs all four rANOVAs. The researcher had certain expectations about the results: greater spatial variability at one kinematic marker in one condition vs. the other (i.e. a pairwise test expectation), and a main effect of block on RT (i.e. an omnibus test comparison). An interaction between condition and kinematic marker is detected.
The researcher decides to do pairwise tests, testing the difference between conditions at each kinematic marker using Tukey's correction for multiple comparisons.
Unfortunately, no effects are detected, not even the predicted effect. The multiple comparison correction may have been too conservative - it often is because it does not account for structural relationships among comparisons (Kruschke, 2014). The researcher

is tempted to just run a single t-test on the predicted difference, with no corrections, as it is noticed that this is the largest difference among all possible differences. Is this a *planned* comparison since the researcher was predicting a difference there anyway? Or, is the researcher running *because* it is the larger difference and *because* the multiple comparison approach did not work. Even more importantly, what if another (not predicted) pairwise difference was noticeably bigger than the others? Would the researcher have run a single test in that instance? If so, the researcher would disproportionately be increasing the type I error rate by only testing the biggest difference in a post-hoc manner.

Moving on, the initially predicted main effect of block on RT is not detected. However, a main effect of block on MT is detected. The researcher reasons that one *should have* expected the effect of block on MT for theoretical reasons. Naturally, the researcher is interested in the specific way in which block affects MT. Would testing pairwise differences be appropriate? Would the researcher have come up with good theoretical reasons to have expected any other unpredicted effect had it been detected? One can see how convoluted multiple testing procedures can get in the frequentist context, particularly when planned tests are not explicitly laid out. Unfortunately, the more that tests are pre-planned, the less flexibility the researcher has to explore interesting hypotheses after collecting the data.

One final point regarding the imaginary scenario is that the researcher did not correct for running multiple ANOVAs. Having not done so, the type I error rate could have been

inflated and the extent to which it may have been is difficult to define without knowing the structural relationship among the effects of different DVs. But, what if having done so would have eliminated the predicted interaction effect? Ultimately, the space of counter factuals make properly running multiple tests in the frequentist context tricky. Upon collecting the data and establishing a reasonable statistical model, the Bayesian method of inference is not dependent on the researcher's testing intentions. The model provides a single posterior distribution which can be marginalized in as many ways that the researcher sees fit (Kruschke, 2014).

Importantly, even with respect to dealing with multiple comparisons, the added value of GPR exceeds the intrinsic benefits of the Bayesian analysis summarized in the previous paragraph. For instance, modelling the entire trajectory and/or its derivatives will naturally minimize the number of (potentially related) DVs that are analyzed in a typical online motor control paradigm. Indeed, standard methods of analysis for identifying visual control mechanisms often fail to take advantage of richness of the full trajectory data set, account for the hierarchical structure of data, and comprehensively estimate all parameters of interest (i.e. using a single statistical model). The thoughtful use of GPRs in the Rstan framework (Stan Development Team, 2017) is a solution to these problems and limitations. The aim of the current thesis was to implement a specific model from the family of Bayesian hierarchical GPR models and to demonstrate the extent to which such a model would benefit researchers in their endeavor to understand the cognitive and perceptual mechanisms underlying human goal-directed reaching.

It is important not to overstate the benefits of Bayesian inference over frequentist inference. Indeed, there is nothing inherently wrong with frequentist statistics when interpreted properly – many of the criticisms brought forth in this section relate to the propensity of researchers to misuse and misinterpret frequentist methods. Also, one should not falsely attribute all the benefits of GPR and hierarchical modelling to Bayesian statistics. For instance, hierarchical models can be implemented quite easily in the frequentist context (for examples, see Bates et al., 2014). GPR can also be implemented in a framework that is not fully Bayesian by finding the maximum likelihood values for the hyperparameters instead of placing priors on them (for foundational work on GPR, see Rasmussen & Williams, 2006).

1.5 STUDY PURPOSE

The purpose of this thesis was to compare the validity of Bayesian hierarchical GPR to that of the standard methods described thus far in inferring visual control processes from human limb trajectory data. Similar methodological endeavors have taken place (for examples, see Khan et al., 2006; and Elliott & Hansen, 2010). However, a novel statistical application is detailed here and was hypothesized to offer stepwise improvements in the validity and comprehensiveness over the standard methods of analysis of human goal-directed movement data.

The thesis purpose was actualized by first running a multitude of common analyses on data collected in a simple goal-directed reaching paradigm for which participant vision was experimentally manipulated. The proposed Bayesian hierarchical GPR analysis was

run on the experimental dataset (Chapter 4) and its results were compared to those of the traditional methods (Chapters 3). In addition, Chapter 5 summarizes the results of a full-scale simulation in which both the proposed GPR method and the gambit of traditional methods were run on a simulated dataset for which the true generative parameter values were known. The purpose of the simulation was to objectively assess the accuracy of the various parameter estimates across all the statistical methods.

It should be emphasized that the purpose was not to address a particular research question in the realm of visual online motor control. Instead, the objective was to propose and evaluate a statistical tool, the benefits of which are applicable to most of the visual online motor control research programme. In fact, many of the benefits, in theory, could even be extended to the field of human limb trajectory research as a whole.

The statistical methods were compared on the basis of their ability to (1) detect a main effect of vision - the existence of which is strongly supported and therefore simply needed to be replicated (Elliott & Hansen, 2010) - and (2) recover the real population parameters from the generative model that was used to the simulated data. In both cases, the depth of information provided about the results (i.e. the parameters of interest) was also weighted. The parameters of interest under different analysis schemes were necessarily be directly comparable (e.g. standard outcome differences will not necessarily be captured by GPR). In many cases, however, the DVs were at least roughly comparable (e.g. R^2 values as a function of proportion of movement vs. population mean function volatility to estimate the extent to which position early on in the movement predicts

position later in the movement). Ultimately, the goal was to propose a candidate best-statistical-tool (GPR) for these purposes and then determine which tool is actually the most *useful* for researchers.

It was hypothesized that effects of visual condition (vision vs. no-vision) would emerge in both the novel and standard methods of analysis, as they have in previous experiments. In addition, it was hypothesized that the proposed Bayesian hierarchical GPR model would be superior to the standard methods in the nuance and accuracy with which it detected experimental effects of interest (i.e., vision vs. no-vision).

CHAPTER 2 METHODOLOGY

2.1 PARTICIPANTS

After piloting, 33 right-handed participants with normal to corrected vision successfully completed the study. They consisted of Dalhousie undergraduate students recruited via Dalhousie's SONA participant pool. This sample size was larger than what is typically collected in the motor control literature and therefore ensured a reasonable likelihood of observing effects. Indeed, Elliott & Hansen (2010), used a sample size of 17 participants and observed effects. After preprocessing, the sample size was reduced to 29 participants. The participants in this usable sample ranged in age from 18 to 22 years, with a median age of 20 years. There were six male and 23 female participants. Informed consent was acquired prior to data collection using the informed consent form presented in Appendix A and following the verbal clarification of the consent form presented in Appendix B. Prior to commencing the experiment, participants read the instructions presented in Appendix C. Participants received 0.5 credits towards one of their university classes for every 30 minutes they spent taking part in the study. The study lasted approximately 30 minutes.

2.2 STUDY DESIGN

Participants completed a single experimental session. There were two within-subject conditions - a vision and a no-vision condition. In the no-vision condition, PLATO

Visual Occlusion Spectacles (Translucent Technologies Inc.) turned from a transparent to translucent state upon movement initiation. Participants completed a total of two experimental blocks, one for each condition. Block order was alternated across participants to counterbalance potential order effects. Each block consisted of 10 practice trials and 20 experimental trials. Although the experiment was self-paced, participants typically did not take long breaks between trials, nor did they between trials.

2.3 PROCEDURE

The experimental procedure was purposively largely similar to that of Elliott & Hansen (2010). Participants made reaching movements to a target that was a 5 $(mm)^2$ white circle placed approximately 31 cm away, along their midline. Each participant had infrared emitting diodes (IREDs) attached to their right index finger with standard medical tape - one on the distant end of their finger and another on the middle part of their finger. The IRED on the distant part of the index finger was the primary marker used for data analysis, with the other marker serving as a backup in cases in which the position of the primary marker was unavailable for multiple samples at a time. The Optotrak 3020 was used to record the position of the IREDs in three dimensions throughout each trial at a sampling rate of 200 Hz (5 ms per sample). The Optotrak 3020 recorded 5000 ms of data for each trial. The recording was manually initiated by the experimenter upon hearing the auditory stimulus signaling that the participant had initiated the trial.

Figure 3 illustrates the sequence of events of a no-vision trial. For both conditions, each trial was initiated once the participant's right index finger touched a square area, the

limits of which were defined by a white fixation cross. The fixation cross had 5 mm arms from its center and therefore covered a 28 mm square area. The cross was placed at the bottom of a 1920 by 1080 pixel HP Pavilion 23tm 23-inch Diagonal Touch Monitor. The touch screen was rotated 90 degrees such that the longer dimension ran along the participant's midline. The target was 31 cm away from the fixation cross, along the participant's midline. Upon touching the fixation cross, a brief auditory stimulus indicated to the participant that the trial had begun. After a random 1000-2000 ms time interval, the target appeared, and participants made reaching movements to it as fast and as accurately as possible. Participants heard the same brief auditory stimulus upon reaching the target. Participants had 3000 ms to reach the target before text appeared on screen indicating that they 'Too Late!'. Conversely, if participants removed their index finger from the fixation cross prior to the presentation of the target, they received a feedback message indicating that they were 'Too Soon!'. In the no-vision condition, participant vision was removed at movement initiation and restored at movement end (when the target was reached). Therefore, participants received terminal feedback regarding the position of the target relative to their limb. They kept their index finger on the target until it disappeared, 3000 ms after the participant had touched it. If participants removed their index finger prior to the removal of the target, they received a feedback message indicating that they needed to keep their finger where it landed. Target disappearance indicated the end of the trial. Participants then returned their finger to the home button to initiate another trial.

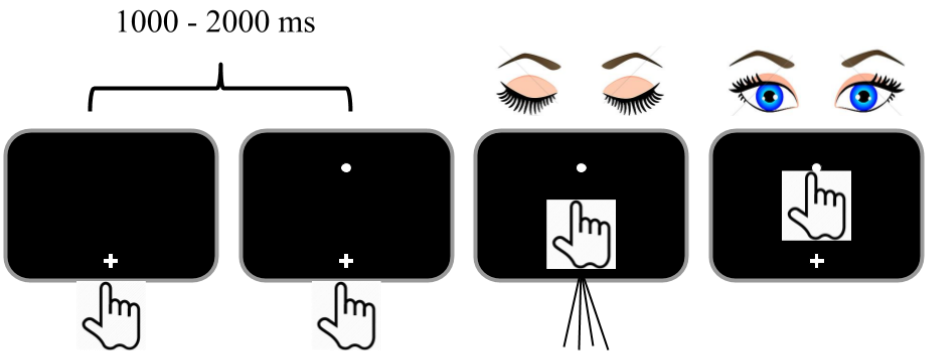


Figure 3 Trial procedure for a no-vision trial. The black squares represent the touch pad. The small white fixation crosses on the black squares represent the fixation cross that needed to be touched to initiate the trial. The small white circles on the black squares represent the target. The lines radiating from the cartoon hand represent movement. The time interval between trial initiation and target presentation varied randomly from trial to trial, between 1000 and 2000 ms. Upon movement initiation, vision was removed using occlusion goggles. This loss of vision is symbolized by the closed cartoon eyes. Participant vision was recovered upon reaching the target. In the full vision condition the only difference was that the participant would not lose their vision upon movement initiation.

2.4 PREPROCESSING

The preprocessing procedure in this section were implemented on the experimental trajectories in each of three dimensions, independently, unless otherwise specified. Section 5.1 describes the way in which the simulated dataset was generated and processed prior to analysis.

2.4.1 Equipment/Experimenter Errors

A custom-made programming script was used to eliminate bad samples (i.e. lost marker) and bad trials (i.e. the participant did not follow instructions or marker was not properly detected throughout the trial). First, trials in which the participant performed the task properly, but in which the touch screen was overly-sensitive or malfunctioned in some other way, were removed. Afterwards, trials in which the Optotrak malfunctioned or was improperly operated by the experimenter were removed. Trials in which the occlusion goggles malfunctioned (likely due to experimenter error) were also removed.

2.4.2 Outlying Samples

For both primary and secondary markers, outlying trajectory samples were identified in order to characterize bad trials for removal from the data set. Outlying samples were defined as those for which the position value in any of three dimensions exceeded upper and lower bound thresholds determined for each dimension via visual inspection. When the Optotrak loses track of a given IRED, the position values for the corresponding samples tend to be unreasonably large and/or variable. If the majority (four or more) of the samples in a 35 ms (seven samples) centered moving average window was outlying, then the data from the corresponding trial was replaced with those of the secondary marker. If secondary marker data also contained a moving average window for which the majority of samples were outlying, then the corresponding trial was removed. There were no trials in which either the primary or the secondary marker had a majority of outlying samples in any moving average window. For certain participants, the preprocessed secondary marker data were used instead of the primary marker data because it was determined via visual inspection prior to preprocessing that the secondary marker data

were definitively cleaner (i.e. there were fewer missing data points). After the completion of the preprocessing steps summarized in this paragraph, each participant had at least 10 trials remaining in each condition (vision and no-vision) suitable for subsequent analysis.

The preprocessing sequence of the previous paragraph did not actually replace any outlying samples – it merely identified bad trials. Therefore, missing value imputation by last observation carried forward was performed to replace the outlying samples. This method replaces each missing value with the most recent present value prior to it. Despite the preprocessing steps explained thus far, four participants still had unusable datasets based on visual inspection of their trajectories and were thus removed. Their trajectories tended to be overly noisy prior to movement onset. It was suspected that when certain participants 'wiggled' their index fingers over the fixation cross to initiate each trial, the Optotrak lost the IRED markers with a greater frequency than usual, therefore resulting in overly noisy data prior to movement initiation.

2.4.3 Filtering

For each of three dimensions, the median position value of the first 500 ms (100 samples) of each trial was subtracted from the entire trial in order to center movement start position at zero. In keeping with the preprocessing methods used by Elliott & Hansen (2010), which are largely in line with what is done in the field (for examples, see Khan et al, 2002; Heath et al., 2004; and Grosbois & Tremblay, 2016), a sixth order low pass forward and reverse butterworth filter with a 8 Hz cutoff was applied to each trial, in each of three dimensions. To get rid of edge effects from the filter, the first 50 ms and last 500

ms of each 5000 ms trial recording were removed.

2.4.4 Differentiation

For each of three dimensions, velocity and acceleration profiles were obtained using forward difference differentiation which involves subtracting the previous value from the current value and dividing that difference by the time interval between samples (5 ms). Subsequently, the first 5 ms (one sample) of each trial recording were removed to eliminate the artifacts of differentiation.

2.4.5 Movement Onset and Offset

For each trial, the custom script identified all the 25 ms (five samples) time intervals for which each sample in a given interval had a corresponding movement amplitude position value greater than 30 mm and a corresponding movement amplitude velocity value greater than 10 mm/s. The first sample in each of those time intervals was considered a potential movement onset sample. The earliest potential movement onset sample was then selected. The sample preceding it and for which the velocity value for movement amplitude changed from negative to positive was identified as the actual movement onset sample. Further, for each trial, the script identified all the 25 ms time intervals for which each sample in a given interval had a corresponding position value greater 270 mm and a corresponding velocity value less than 100 mm/s. The first sample in each of those time intervals was considered a potential movement offset sample. The earliest potential movement offset sample was then selected. The sample which succeeded it and for which

the velocity value changed from positive to negative (i.e. at which the movement stopped) was identified as the actual movement offset sample. Trials for which either movement onset or offset was unidentified were discarded. A total of 18 trials, spread across five participants, were discarded during this procedure. Importantly, each participant still had at least 10 usable trials (out of 20 completed trials) for each experimental condition after this preprocessing step for subsequent analysis. The trajectory data were then trimmed so that each trial started at movement onset and ended at movement offset.

2.4.6 Normalization

Trial-by-trial data were normalized with respect to time such that the time points ranged from zero to one and were equally spaced. The trajectory data were then downsampled to allow for the estimation of sensible participant-wise (averaging over trials) and population-level (averaging over participants) trajectories, as not all reaching movements lasted the exact same amount of time. In principle, downsampling had no effect on the *proportional* TAPV results since this DV was naturally normalized with respect to time. For the discontinuity, R^2, spatial variability profile, and epsilloid analyses, time points were broken down into 21 equally spaced bins so that the normalize time length of each bin was 0.05. For each of the three dimensions and for each trial, the mean position value in each time bin was then obtained.

2.4.7 Scaling

The following preprocessing procedure was only applied for the Bayesian analysis. The trajectories were scaled with respect to both position and time. In the former case, all the position data were simultaneously scaled to have a mean of zero and a standard deviation of one (this was *not* done on a condition-, trial-, or participant-wise basis). The time dimension was scaled so as to range from zero to one (this *was* done on a trial-wise basis). These normalization steps will be taken to facilitate the specification of uninformative priors and the mean function of the GPs (a vector of zeros). The parameter estimates of the Bayesian hierarchical GPR model were interpreted in the scaled space (i.e. they were *not* back-transformed to their original scales). In principle, this did not impact the interpretation of the results as the focus was on the differences between the estimates of the two experimental conditions and the timing of those differences along the proportion of movement axis.

2.5 DEFINING VARIOUS MEASURES OF VARIABILITY

The present thesis introduces multiple terms that relate to some sort of movement variability measurement. For clarity, these terms will be explicitly defined. The spatial variability profile of a given participant is simply a measure of the variability (SD) of the trajectory across trials at each normalized time point of the movement. As such, the spatial variability profile represents the trial-by-trial or trial-wise variability of a set of movement trajectories for a particular participant. An average of participant spatial variability profiles is obtained to make inferences about the population spatial variability profile, for each condition. Likewise, in the context of the Bayesian hierarchical GPR model, the participant and population noise functions correspond to the participant and

population spatial variability profiles. The term "function" was used to highlight that these estimates of trial-by-trial variability are modelled as functions, sampled from GPs. The term "noise" refers to the fact the functions are measures of variability and thus represent noise either among trials (participant noise functions) or among participants (population noise functions).

The hyperparameters of the Bayesian hierarchical GPR model also include certain measures of variability. Namely, both the population noise and mean functions have corresponding amplitude and volatility variability hyperparameters. These variability hyperparameters correspond to the variability (SD) of the amplitude and volatility hyperparameters across participants, for both the population noise and mean functions. For example, if the population volatility variability is large, it means that there is a lot of variability among the volatilities of the participant mean functions (i.e. some participants have very 'wiggly' mean functions whereas others have more linear mean functions).

CHAPTER 3 TRADIONAL ANALYSES

3.1 METHODS

The statistical analyses presented in this chapter were performed in R (2016) and were run on experimental data (collected from the experiment described in subsection 2.3). The following statistical analyses were run: rANOVAs[2] on proportional TAPV, discontinuities, R^2, and standard outcome differences (RT, MT, CE, VE) (for comparison among these methods and subsequent discussion, see Elliott & Hansen, 2010). The focus was on the R^2, variability profile, and spline fitting analyses as these were most stringent comparisons to the proposed Bayesian method, but the remaining rANOVA analyses were run to compare the results to previous work (e.g. Elliott & Hansen, 2010). Only significant effects are reported. In order to compare effect sizes among the measures listed above within the current study, generalized eta squared (η_G^2) effect sizes were reported (for discussion regarding effect sizes, see Olejnik & Algina, 2003). The ez package in R was used to run the rANOVA and effect size analyses, and to generate the corresponding graphs. Fisher's Least Significant Differences (FLSDs) were plotted to evaluate the variability associated with within-subject effects.

Table 1 summarizes the predicted experimental effect of vision (vision minus no-vision

[2] Some of the rANOVA analyses, as will be described later in the chapter, were one-way rANOVAs and therefore effectively t-tests. In these cases the only factor was vision condition with two levels: vision and no-vision.

condition) for each DV. These predictions were based on the results obtained from Elliott & Hansen (2010). Also, these predictions could have largely been derived from the literature cited throughout Chapter 1.

DV **Analysis method**	**Expected effect (vision minus no-vision)**
Proportional TAPV	Positive: greater proportional amount of time spent after peak velocity with vision
Discontinuities	Positive: more discontinuities in kinematic profiles with vision
R^2	Negative: less variability explained at PV and PD with vision
RT	Negative: shorter reaction time when vision is expected
MT	Null
CE	Negative: movements are more accurate with vision
VE	Negative: movement error is less variable with vision
Variability profile	Positive: greater amount of trial-wise variability with vision around PD
Epsilloid method	Positive: greater amount of trial-wise variability with vision mid-way through the movement
fANOVA	Unknown

Table 1 Expected experimental effects (the value of the DV in the no-vision condition subtracted from the value of the DV in the vision condition) for each DV under investigation. Expected interaction effects between vision condition and proportion of movement are also stated for the variability profile and epsilloid methods (DV: trial-wise variability), and the fANOVA analysis.

3.1.1 Proportional TAPV

For each of three dimensions, for each participant, and for each trial, the time point corresponding to the peak velocity value was obtained. The mean of those time points over all trials was then calculated for each condition to obtain participant-wise time before peak velocity values. These values were then subtracted from one to obtain TAPV values for each participant and each condition. A one-way rANOVA, with vision condition as a factor, was run on the TAPV values for movement amplitude.

3.1.2 Discontinuities

For each of three dimensions, for each participant, the trials in which there were either negative velocity or negative acceleration values were identified as trials with discontinuities. The number of trials in which there was a discontinuity was then calculated for each condition to obtain participant-wise discontinuity counts. A one-way rANOVA, with vision condition as a factor, was run on the discontinuity counts for movement amplitude.

3.1.3 R^2

For each of three dimensions, the position value at the time of peak acceleration (PA),

time of peak velocity (PV), time of peak deceleration (PD), and end of movement (END) for each trial was obtained. Univariate linear regression was performed for each participant and each condition by predicting the position value at the end of movement from the position values at each kinematic marker across trials. Condition-wise R^2 values for each participant were then obtained from the linear regression results. A 2 (vision condition) by 4 (kinematic marker: PA, PV, PD, END) rANOVA was run on the R^2 values for movement amplitude.

3.1.4 Outcome Measures

RTs, MTs, and response positions with respect to the touch screen for each trial were computed in real time via the experiment Matlab script. RT was the time between the target appearing and the participant lifting their figure from the touch screen. MT was the time between the participant lifting their finger from the start button and contacting the target.

Target error was defined simply as the absolute distance between the response position and the target position. CE (amplitude) was defined as the distance between the response position and the target position for movement amplitude, whereas CE (direction) was defined similarly but for movement direction. The participant-wise CEs (amplitude and direction) were calculated by taking the mean over the CEs (amplitude and direction, respectively) for all trials in each condition. Participant-wise VEs (amplitude and

direction) were calculated by taking the standard deviation of CEs over all trials in each condition. A one-way rANOVA, with vision condition as a factor, was run on each outcome measure.

3.1.5 Spatial Variability Profile

For each of three dimensions, the participant-wise and trial-wise standard deviations of the position values at each sample of the normalized trajectory data were obtained. The mean of the standard deviations at each sample over all trials was calculated for each condition to obtain participant-wise spatial variability profiles. For each of three dimensions, the spatial variability at the time of PA, time of PV, time of PD, and END for each trial was obtained. A 2 (vision condition) by 4 (kinematic marker: PA, PV, PD, END) rANOVA was run on the spatial variability values for movement amplitude. FLSDs were plotted to evaluate the variability associated with within-subject effects.

3.1.6 Epsilloid Method

The product of the spatial variability profiles (see subsection 3.1.5) for each participant in each dimension at each time point was taken and multiplied by four thirds of pi to obtain epsilloid volumes. These values were subject to a 2 (vision condition) by 21 (proportion of movement) rANOVA. FLSDs were plotted to evaluate the variability associated with within-subject effects.

3.1.7 fANOVA

The partially preprocessed trajectory data (see preprocessing steps leading up to subsection 2.4.6) were normalized according to the procedures described by Gallivan & Chapman (2014). Their custom MATLAB scripts were used to normalize the trajectory data with respect to time and run the fANOVAs[3]. Specifically, order six splines were fit to the position values of each of the three dimensions, with a spline for every sample. The trajectory data were then smoothed using a roughness penalty on the fourth derivative. The result was a mathematical definition of each the trajectory data in each dimension across time. The position values in each dimension were sampled from 200 normalized time points, equally-spaced within each trial. Participant mean trajectories were then obtained for each dimension by taking the mean of the position values over all trials for each time point and for each condition. Since there were only two within-subject conditions, the fANOVAs were reduced to a functional-t-tests. A functional-t-test amounts to running pairwise t-tests on the condition mean trajectories at each time point, for each dimension. FLSDs were plotted to evaluate the pair-wise differences between the two vision conditions at each time point.

3.2 RESULTS & DISCUSSION

[3] For MATLAB scripts, see http://www.per.ualberta.ca/acelab/.

3.2.1 Proportional TAPV

The estimates of proportional TAPV as a function of vision condition are presented in Figure 4. The TAPV analysis revealed a main effect of vision condition [$F(1,28) = 13.5$, $p = 0.001$, $\eta_G^2 = 0.079$]. TAPV was greater in the vision condition than it was in the no-vision condition. The direction of the effect was consistent with what was observed by Elliott & Hansen (2010). Along these lines, Hansen et al. (2006) observed greater absolute TAPVs when participants had knowledge, prior to movement onset, that vision would be available during the movement[4]. In line with the model proposed by Elliott et al. (2017), participants may have adopted a strategy of maximizing the potential for online visual control when they knew that vision would be available.

[4] The authors also observed greater MTs when participants had prior knowledge of vision availability. Therefore, in their case, participants seemed to have compromised overall movement speed in order to make greater use of vision during the latter phase of the movement.

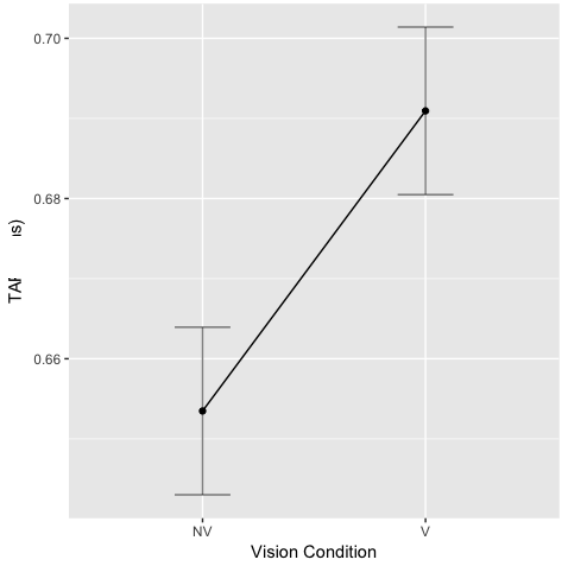


Figure 4 Proportional TAPV plotted as a function of vision condition (x-axis). The no-vision and vision conditions are represented by "NV" and "V", respectively. Error bars represent FLSDs.

3.2.2 Discontinuities

The estimates of discontinuity count as a function of vision condition are presented in Figure 5. The discontinuity analysis revealed a main effect of vision condition [$F(1,28) = 6.08$, $p = 0.020$, $\eta_G^2 = 0.055$]. The discontinuity count was greater in the vision condition than it was in the no-vision condition. The main effect of vision was consistent with what was observed by Elliott & Hansen (2010). With access to visual feedback regarding their limb and the target throughout the movement and with prior knowledge about that access, participants appeared to have made more discrete corrections in the vision condition (for discussion regarding limb-target phase of movement, see Elliott et al., 2017).

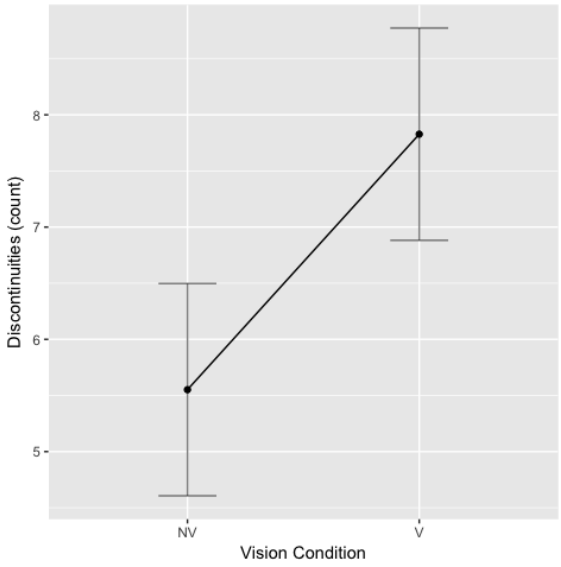


Figure 5 Discontinuity counts plotted as a function of vision condition (x-axis). The no-vision and vision conditions are represented by "NV" and "V", respectively. Error bars represent FLSDs.

3.2.3 R^2

The estimates of R^2 as a function of vision condition and kinematic marker are presented in Figure 6. The R^2 analysis revealed a main effect of vision condition [$F(1,28) = 6.44$, $p = 0.017$, $\eta_G^2 = 0.037$]. The R^2 values were greater in the no-vision condition than they were in the vision condition. Overall, the R^2 values were low (between 0.05 and 0.15) which is consistent with the grand mean of 0.13 observed by Elliott & Hansen (2010)[5]. Although the interaction effect was not significant [$F(2,56) = 1.93$, $p = 0.15$, $\eta_G^2 = 0.024$],

[5] These authors did not observe an effect of vision on R^2 however.

the overall *trend* of the data in which the R^2 values were greater in the no-vision condition, particularly at PD, was consistent with the findings of Heath et al. (2004)[6]. The reduction in the linear relationship, in the vision condition, between the position of the limb later in the movement and at the end of the movement may have resulted from participants making a greater number of online corrections later in the movement when vision was available. There is no way to discriminate, however, whether the online corrections corresponded to the impulse control or limb-target (discrete corrections) phase of the movements (Elliott et al., 2017).

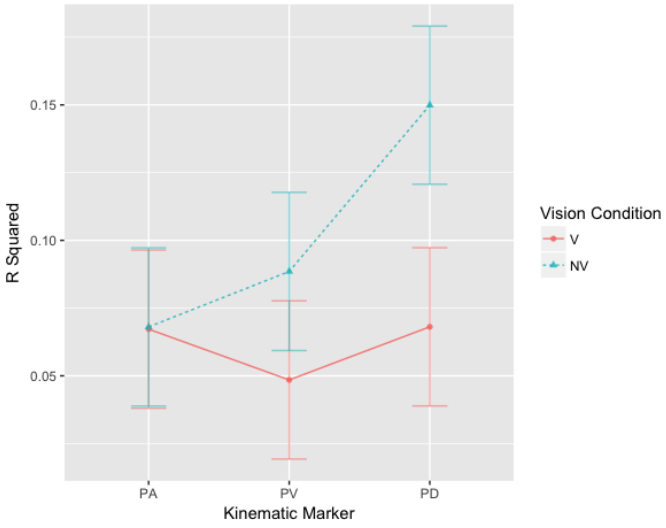


Figure 6 R^2 values plotted as a function of kinematic marker (x-axis) and vision condition (line/point color and type). The vision and no-vision conditions are represented by "V" and "NV", respectively. Error bars represent FLSDs.

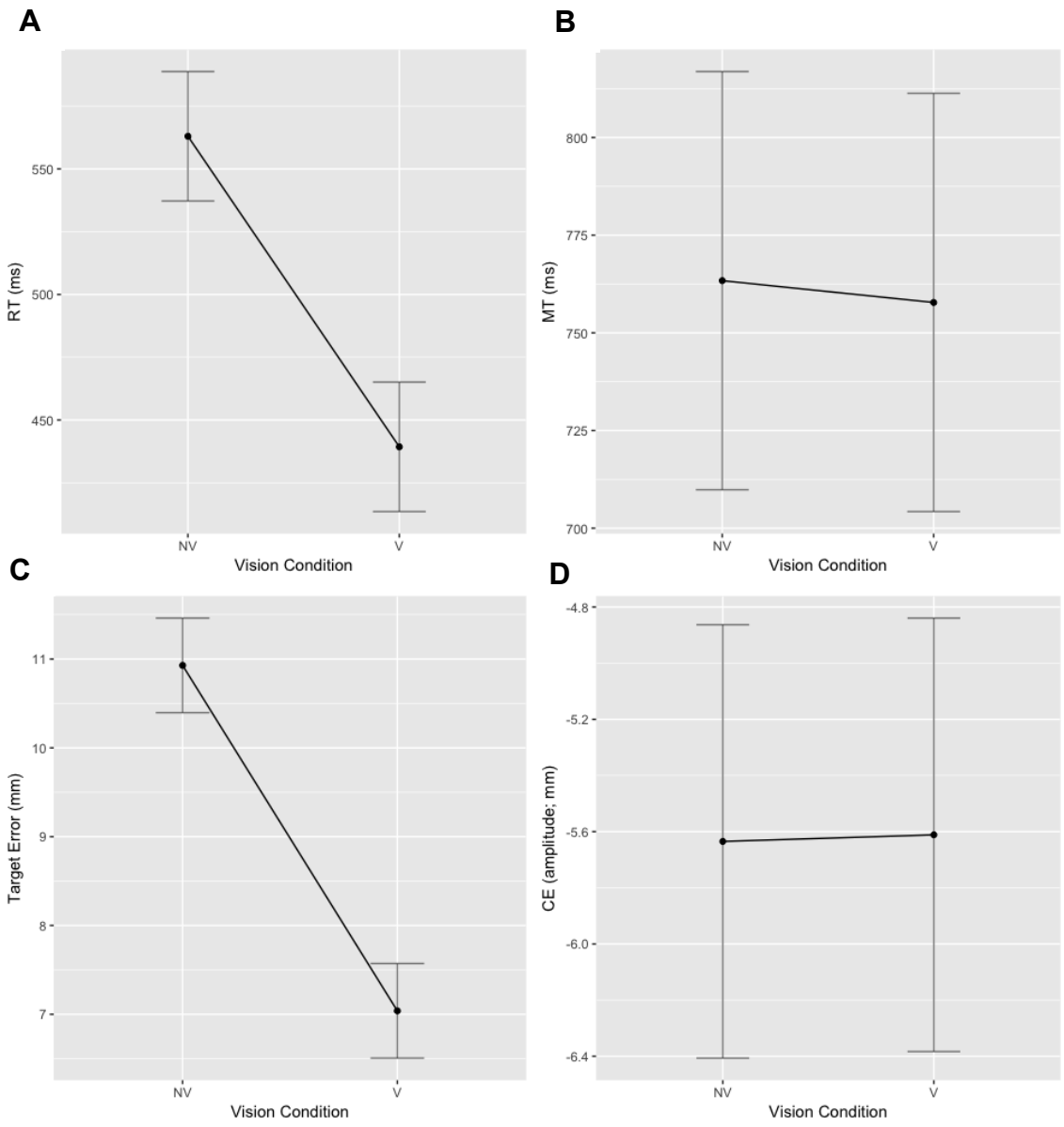
[6] Although the trend was similar, these authors observed much larger R^2 values overall.

3.2.4 Outcome Measures

The estimates of each outcome measure as a function of vision condition are presented in Figure 7. The RT analysis revealed a main effect of vision condition [$F(1,28) = 24.1$, $p < 0.001$, $\eta_G^2 = 0.121$]. RTs were greater in the no-vision condition than they were in the vision condition, a finding consistent with that of Hansen et al. (2006). There were no significant effects of MT, a finding consistent with that of Elliott & Hansen (2010). The target error analysis revealed a main effect of vision condition [$F(1,28) = 56.0$, $p < 0.001$, $\eta_G^2 = 0.424$].

Target error was greater in the no-vision condition than it was in the vision condition [$F(1,28) = 56.0$, $p < 0.001$, $\eta_G^2 = 0.424$], a finding consistent with the notion that movement accuracy should be greater with vision. The VE (amplitude) analysis revealed a main effect of vision condition [$F(1,28) = 114.9$, $p < 0.001$, $\eta_G^2 = 0.583$]. The VE (amplitude) values were greater in the no-vision condition than they were in the vision condition. This effect was consistent with that of Elliott & Hansen (2010) in both directionality and magnitude: the authors observed an average VE (amplitude) error of 8.04 mm for the no-vision condition, and of 4.06 mm for the vision condition. The VE (direction) analysis revealed a main effect of vision condition [$F(1,28) = 93.4$, $p < 0.001$, $\eta_G^2 = 0.536$]. The VE (direction) values were greater in the no-vision condition than they were in the vision condition, a finding consistent with that of Elliott & Hansen (2010).

The CE (amplitude and direction) analyses failed to reveal any significant effects. The null effect of CE (amplitude) was not predicted as one would have expected participants to undershoot the target more so in the no-vision condition (i.e. participants should have been more accurate in the vision condition). Indeed, the model posited by Elliott et al. (2017) predicts that there would be less undershooting in the vision condition as there was less spatial variability at the end of the movement in this condition (see Figures 9, 10, and 19, to be discussed later). Perhaps the difference in spatial variability at the end of the movement was not sufficiently pronounced (as might be inferred by Figure 8) to generate the corresponding CE (amplitude) effect.


A
RT (ms)
550
500
450
NV
V
Vision Condition
B
MT (ms)
800
775
750
725
700
NV
V
Vision Condition
C
Target Error (mm)
11
10
9
8
7
NV
V
Vision Condition
D
CE (amplitude; mm)
-4.8
-5.2
-5.6
-6.0
-6.4
NV
V
Vision Condition

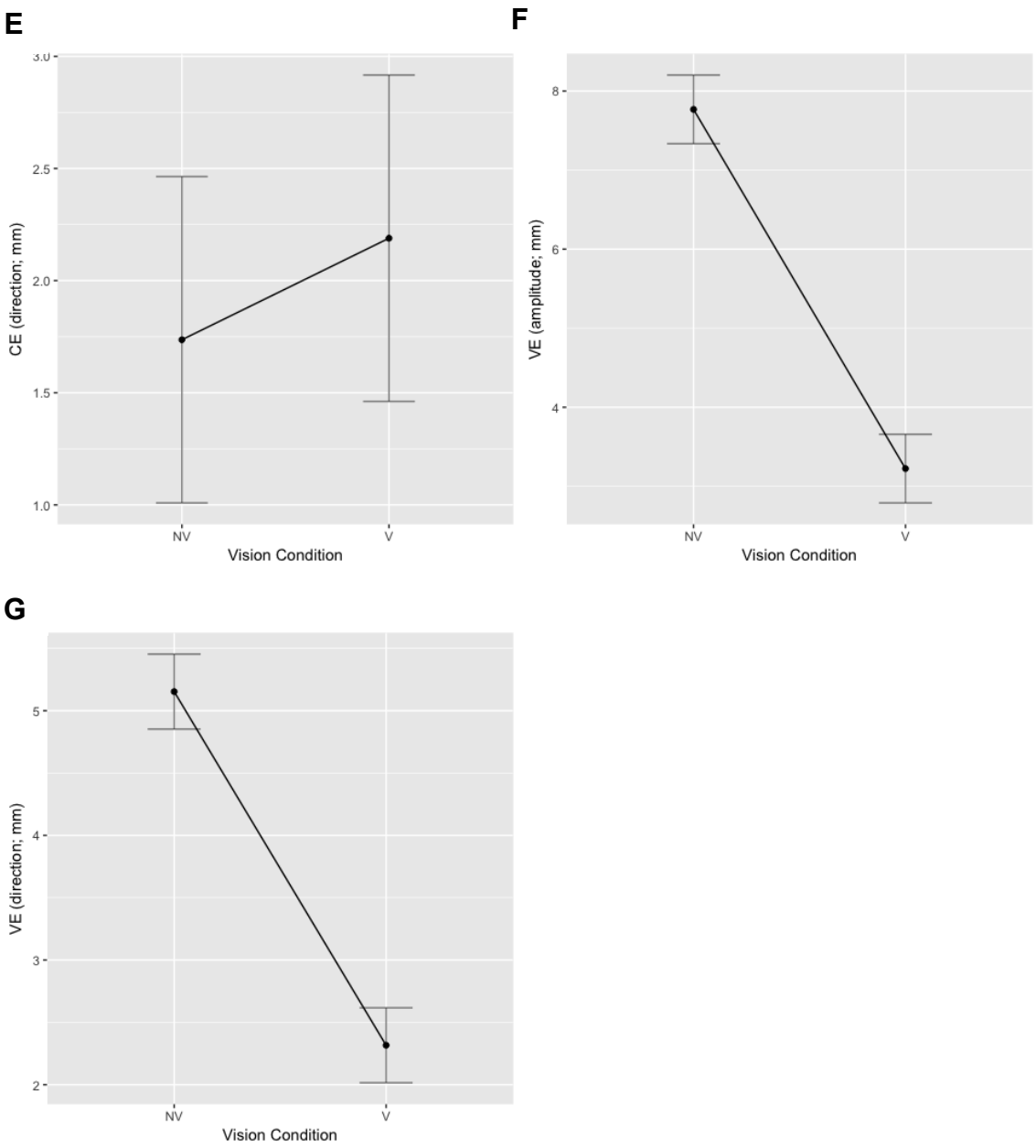


Figure 7 RT (ms; A), MT (ms; B), target error (mm; C), CE (amplitude (D) and direction (E) in mm), and VE (amplitude (F) and direction (G) in mm) are plotted as a function of vision condition (x-axis). The no-vision and vision conditions are represented by "NV" and "V", respectively. Error bars represent FLSDs.

3.2.5 Spatial Variability Profile

The estimates of spatial variability as a function of vision condition and kinematic marker are presented in Figure 8. The spatial variability profile analysis revealed a main effect of kinematic marker [$F(3,84) = 99.2$, $p < 0.001$, $\eta_G^2 = 0.476$]. The analysis also revealed a vision condition by kinematic marker interaction effect [$F(3,84) = 7.03$, $p < 0.001$, $\eta_G^2 = 0.024$]. The interaction effect was characterized by greater spatial variability in the vision condition compared to the no-vision condition early in the movement, with a reversal in that trend at the end of the movement. However, Macuhly's test for sphericity was violated for both the main effect of kinematic marker [$W = 0.273$, $p < 0.001$] and the interaction effect [$W = 0.368$, $p < 0.001$]. Using Greenhouse-Geisser corrections, the main effect of kinematic marker [$\varepsilon = 0.543$, $p < 0.001$] and the interaction effect [$\varepsilon = 0.606$, $p = 0.003$] remained.

The interaction effect was consistent with what was observed by Elliott & Hansen (2010). The early increase in spatial variability with vision may be attributable the inherent increase in movement variability associated with guiding the limb more rapidly to the target under full vision conditions (see proportional TAPV results in subsection 3.2.1). It could also be a consequence of online corrections made during the impulse control phase of the movement (Elliott et al., 2017). Conversely, the decrease in spatial variability at the end of the movement is likely a manifestation of the advantage of having the availability of visual feedback during the movement with respect to movement accuracy,

perhaps attributable to both the impulse control and the limb-target regulation phases of the movement.

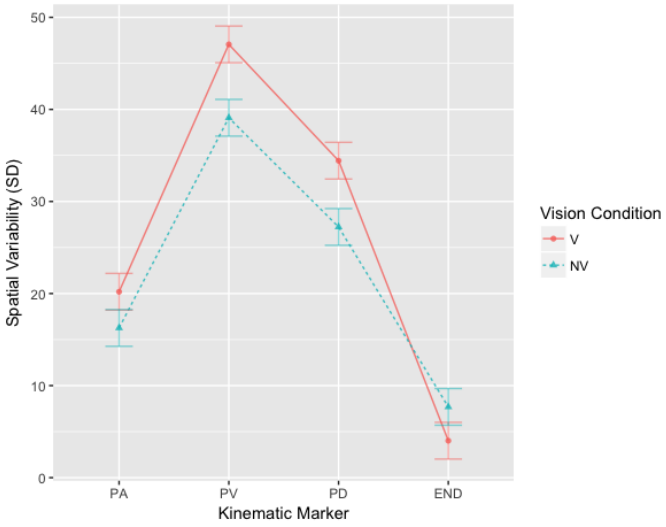


Figure 8 Spatial variability (SD) values plotted as a function of kinematic marker (x-axis) and vision condition (line/point color and type). The vision and no-vision conditions are represented by "V" and "NV", respectively. Error bars represent FLSDs.

3.2.6 Ellipsoid Method

The estimates of epsilloid volume (3D spatial variability) as a function of vision condition and proportion of movement are presented in Figure 9. The ellipsoid volume analysis revealed a main effect of proportion of movement [$F(20,560) = 36.7$, $p < 0.001$, $\eta_G^2 = 0.374$]. The analysis also revealed a vision condition by proportion of movement effect [$F(20,560) = 5.06$, $p < 0.001$, $\eta_G^2 = 0.030$]. The interaction effect was

characterized by greater ellipsoid volumes (spatial variability) in the vision condition compared to the no-vision condition early in the movement, with a reversal in that trend at the end of the movement. However, Macuhly's test for sphericity was violated for both the main effect of proportion of movement [$W < 0.001$, $p < 0.001$] and the interaction effect [$W < 0.001$, $p < 0.001$]. Using Greenhouse-Geisser corrections, the main effect of kinematic marker [$\varepsilon = 0.072$, $p < 0.001$] and the interaction effect [$\varepsilon = 0.097$, $p = 0.013$] remained.

The interaction effect was consistent with what was observed by Elliott & Hansen (2010). It can be seen that the epsilloid method and the standard spatial variability profile analysis offer similar insights into the nature of the movement trajectories (see Figure 8 above). Ultimately, both analysis methods measure the variability of the movement across trials and through time. However, the epsilloid method provides a more nuanced and complete (3D vs. 1D) perspective of the trial-wise variability of the movements than does the standard spatial variability profile analysis. Conversely, the standard spatial variability profile analysis may offer the benefit of explicitly evaluating the trial-wise variability of movements at specific, theoretically meaningful kinematic markers.

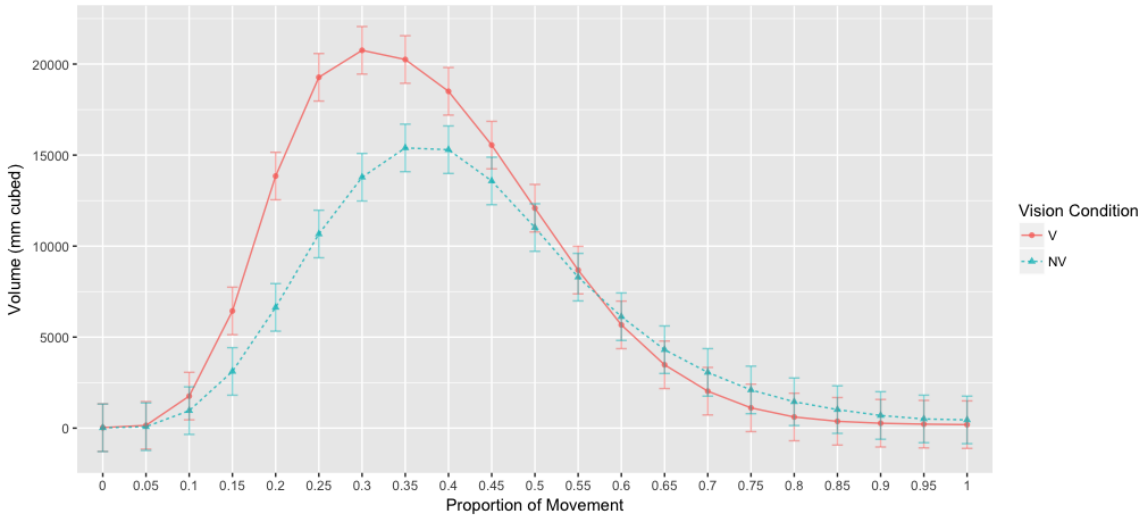


Figure 9 Epsilloid volume ($(mm)^3$) values plotted as a function of proportion of movement (x-axis) and vision condition (line/point color and type). The vision and no-vision conditions are represented by "V" and "NV", respectively. Error bars represent FLSDs.

3.2.7 Overall rANOVA Results

The expected experimental effects largely matched the actual main effects of vision condition (Table 2). The only DV for which the actual main effect of vision was not accurately predicted was CE (amplitude and direction). The main effects on CE were null. The largest effect sizes (greater than 0.5) were attributable to VE (amplitude and direction). The directionality of the effects was largely consistent with that observed by Elliott & Hansen (2010). Further, in terms of evaluating the main effect of vision, the present study and that of Elliott & Hansen (2010) both observed that VE (amplitude and direction) generated the largest effect sizes. Overall, these results suggest that prior knowledge regarding the access to vision (1) permits participants to take less time to plan

their movements, (2) incentivizes participants to maximize their access to visual feedback by getting their limb in the vicinity of the target more rapidly, and (3) results in a greater number of online corrections (discrete or continuous) (for broader discussion on these phenomenon, see Elliott et al, 2017).

DV	**Expected effect (vision minus no-vision)**	**η_G^2 Main effect of vision**
Proportional TAPV	Positive	0.079
Discontinuities	Positive	0.055
R^2	Negative	0.037
RT	Negative	0.121
MT	Null	Null
Target Error	Negative	0.424
CE (amplitude, direction)	Negative	Null
VE (amplitude)	Negative	0.583
VE (direction)	Negative	0.536

Table 2 Expected experimental effects (the value of the DV in the no-vision condition subtracted from the value of the DV in the vision condition) for several DVs. The actual effect sizes of the main effects of vision are listed. If the direction of the expected and actual effects were consistent with one another, the corresponding rows were highlighted in green. If the predicted and actual effects did not match each other, the corresponding rows were highlighted in red.

Both the variability profile analysis and the epsilloid analysis produced the expected results (Table 3). The effect size of the interaction between vision condition and proportion of movement was slightly greater for the epsilloid method. Both variability analyses were sufficiently powerful to detect effects of vision condition over the course of the movement trajectories thereby allowing the researcher to determine where the vision condition effects emerged temporally. Overall, the trial-by-trial variability analyses appear useful in detecting the early effects of vision on the frequency of online corrections and/or movement planning (faster movements tend to generate more variability), and the latter effects of vision on movement accuracy and precision (to put results in context of multiple process model of goal-directed reaching, see Elliott et al., 2017).

Analysis method	Expected effect (vision minus no-vision mid-movement)	η_G^2 Interaction effect (vision condition by proportion of movement)
Variability profile	Positive	0.024
Epsilloid method	Positive	0.030

Table 3 Expected experimental interaction effects (the spatial variability values in the no-vision condition subtracted from the spatial variability values in the vision condition as a function of proportion of movement) for two analysis methods. The actual effect sizes of the interaction effects of vision condition by proportion of movement are listed. If the expected and actual effects were consistent with one another the corresponding rows were highlighted in green. If the predicted and actual effects did not match with one another, the corresponding rows were highlighted in red.

3.2.8 fANOVA

The population mean trajectories for movement amplitude differed substantially between conditions such that participants covered the distance required to reach the target earlier in their proportional movements in the vision condition than they did in the no-vision condition (Figure 10A). The population mean trajectories for movement direction also differed substantially between conditions (Figure 10B). The pattern of results is less clear in this case, but it appears as though in the vision condition participants deviated from their midline earlier in their proportional movements than they did in the no-vision condition. The population mean trajectories along the vertical movement axis also differed substantially between conditions such that, midway in their movements,

participants brought their hands further up in the no-vision condition than they did in the vision condition (Figure 10).

The findings for movement amplitude are consistent with those of the proportional TAPV analysis. Participants covered more ground for a given proportion of movement if they were aware that vision would be available (Figure 10A). Consequently, participants spent more time after peak velocity in the vision condition. These findings are consistent with the notion that knowledge regarding the availability of vision increases the speed at which participants move their limb within the vicinity of the target in order to maximize the benefit of online visual feedback (Elliott et al., 2017). The results of movement direction corroborated this argument as the no-vision condition mean trajectory appears to be slightly phase-shifted to the right of the vision condition mean trajectory (i.e. in the early phases of the movement, the natural movement path for movement direction is traversed more rapidly in the vision condition). This finding appears to simply corroborate the result for movement amplitude. It should be emphasized that although the fANOVA technique was not explicitly designed to compare trajectories between vision conditions, it does capture the results from TAPV analysis while providing a more nuanced representation of the results.

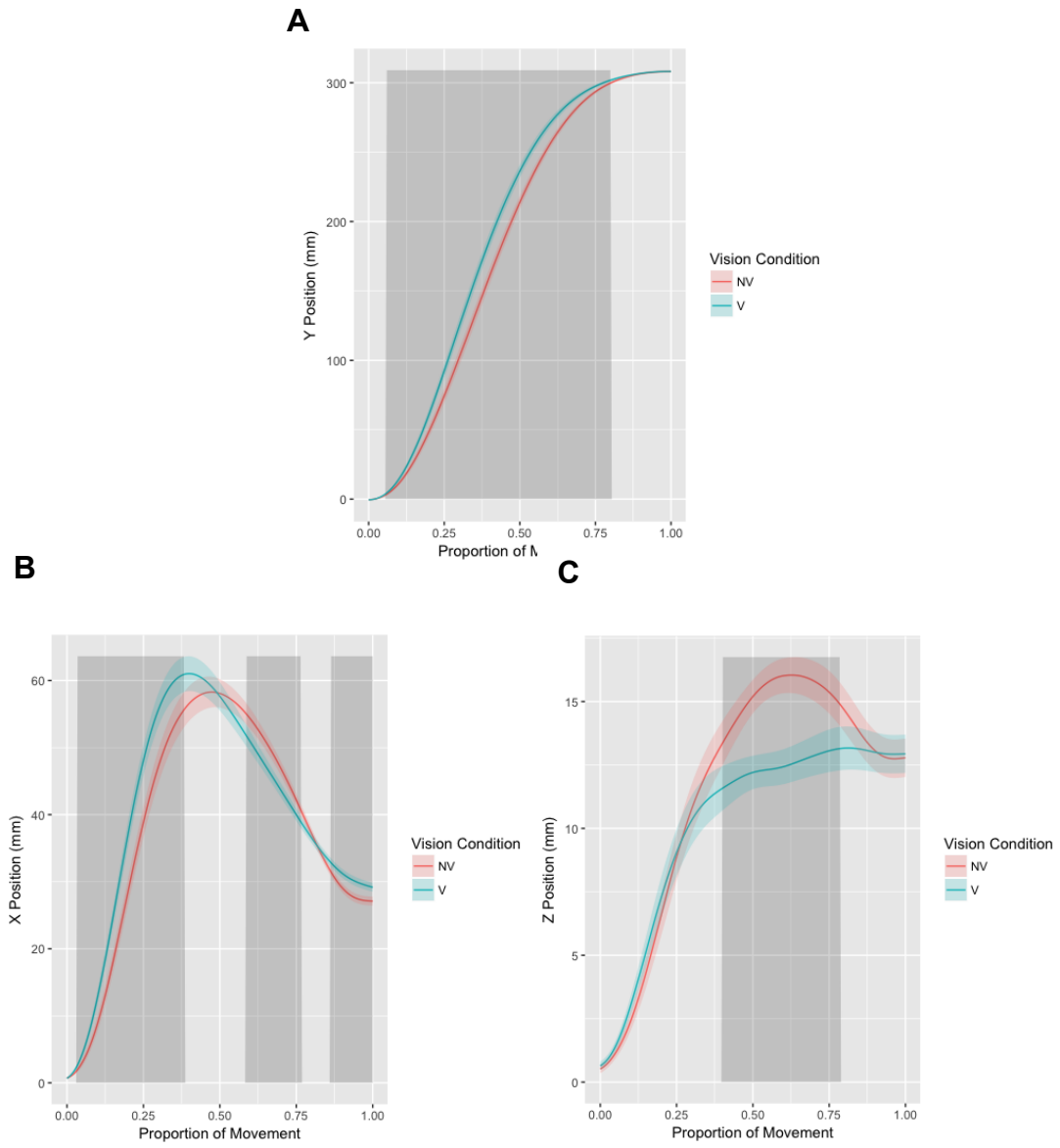


Figure 10 Position values (mm) plotted as a function of proportion of movement (x-axis) and vision condition (line/point color and type) for movement amplitude (y coordinate; A), movement direction (x coordinate; B), and along the vertical (z coordinate) movement axis (C). The vision and no-vision conditions are represented by "V" and "NV", respectively. The gray shaded areas represent the time intervals in which the differences between vision conditions were statistically significant. Error bands represent FLSDs, one for each time point. If the error bands failed to overlap at any time point, then the difference at that time point was statistically significant according to an uncorrected t-test.

CHAPTER 4 BAYESIAN HIERARCHICAL GPR

4.1 METHODS

The statistical analyses presented in this chapter were performed in R (2016) and were run on experimental data (collected from the experiment described in subsection 2.3). GPR was directly compared to the R^2 analysis, the variability profile analysis described in Elliott & Hansen (2010), and the spine fitting method described in Gallivan & Chapman (2014).

The trajectory data for the Bayesian hierarchical GPR analysis were subjected to the preprocessing steps described in subsection 2.4, except for the normalization procedure presented in subsection 2.4.6. For the GPR analysis, the time points were broken down into 15 equally spaced bins so that the normalize time length of each bin was approximately 0.071. The GPR analysis required greater downsampling than the R^2 or spatial variability profile analyses in order to make the estimates of the trajectories computationally tractable[7]. For each of the three dimensions and for each trial, the mean position value in each time bin was then obtained.

[7] *The number of samples that go into the trajectories that are modelled determines the size of the autocorrelation matrix, which is the primary computational bottleneck of the analysis described in the subsection 4.1.*

A technical description of the GPR model that was used is described below. Limb positions for a given trial were sampled from a normal distribution with a mean and standard deviation, a measure of trial-wise variability, specified by participant-specific functions. The participant-specific mean and trial-wise variability functions were calculated as the sum of two GPs: a population GP and a participant-level deviation GP with hyperparameters specific to a given participant. The participant-level deviation functions represented the deviation of a particular participant's function from the population-level function. The participant-level deviation mean and trial-wise variability hyperparameters were sampled from cauchy (volatility hyperparameter) and normal (amplitude hyperparameter) distributions that were centered at zero and had standard deviations with weakly informative weibull priors. The population mean and trial-wise variability GPs were defined by volatility and amplitude hyperparameters with weakly informative cauchy and weibull priors respectively. All GPs in the model were specified by uniform zero mean functions and a exponentiated quadratic function covariance functions. The posterior distribution was sampled using Markov Chain Monte Carlo (MCMC) methods outlined in RStan (Stan Development Team, 2017). The model structure described in this paragraph was applied to each condition, with the parameters estimated simultaneously. The structure of the model is represented schematically in Figure 11. Finally, the model was run for the trajectory data for movement amplitude.

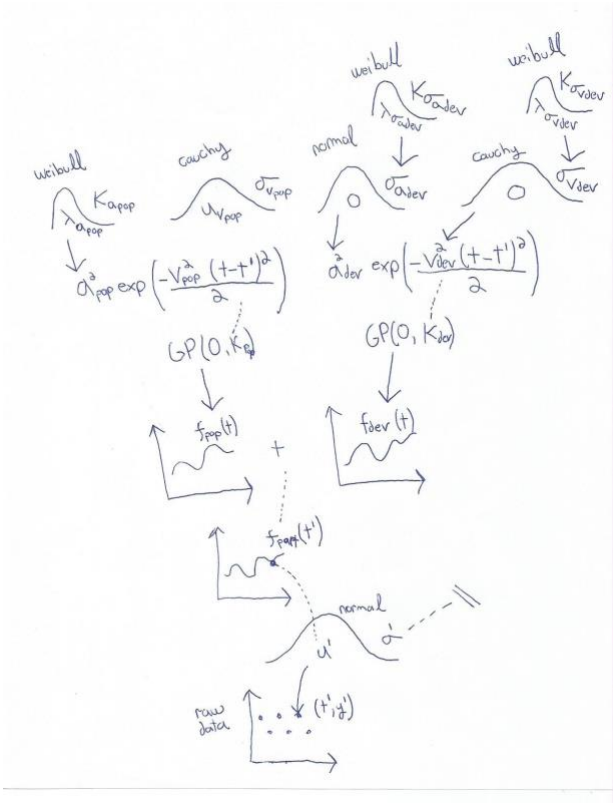


Figure 11 Diagram of Bayesian hierarchical GPR model structure. From the bottom to the top, the raw data were modelled as having been generated from a function with a GP prior. The covariance kernel was an exponentiated quadratic for which the hyperparameters were given priors. Broadly, arrows indicate that a particular parameter or data point was modelled as having been sampled from the distribution from which the arrow originates, with the exception of priors which did not exactly have this interpretation. Dotted lines represent mathematical equality between different representations of the same parameter. The subscripts of the parameters have the following meanings: 'part' = participant; 'pop' = population; 'dev' participant-level deviation. The double-bars, which appear once at the end of a dashed line originating from the bottom normal distribution's sigma parameter, indicate that the logarithm of the standard deviation represented by the sigma symbol had the same model structure as the mean represented by the mu symbol. In other words the participant-specific mean and trial-wise variability functions (in log-space) were modelled identically.

The following validation procedures were run on the model for both the real and the simulated datasets. The model was validated with respect to the convergence its Markov

chains (for discussion, see Gelman et al., 2014; and Stan Development Team, 2017). The potential scale reduction statistic (Rhat) was intended to be as close as possible to 1.00. The number of effective samples for each parameter estimate was intended to exceed 1000. These metrics indicated the extent to which the posterior distribution was properly sampled. The procedure of running a full-scale simulation by generating data from the Bayesian hierarchical GPR model and then running the model on those simulated data by its very nature constitutes a method of validating the analytical model.

Before presenting the results of the Bayesian analysis, it is important to clarify how these results are to be properly interpreted. Bayesian parameter estimates indicate the probability that the value of true population parameter under examination falls within the estimate. The proper interpretation of Bayesian parameter estimates is often erroneously applied to frequentist parameter estimates (for discussion regarding proper interpretation of CIs, see Cumming & Finch, 2005; also, for discussion regarding distinction between Bayesian and frequentist interpretations of uncertainty, refer to section 1.4).

Specifically, the primary inferential tool of the Bayesian method, the posterior distribution, indicates the relative credibility of parameter values among an infinite set of possible values (Kruschke, 2014). The information contained in the posteriors are summarized by highest density intervals (HDIs), a particular type of credible interval (for introduction to credible intervals, refer to section 1.4). An HDI is defined as the interval that captures 95% of the posterior distribution density and the bounds of which

correspond to the greatest probability density values. One can state about a 95% credible interval that there is 95% probability that the true population parameter value falls within the range of values bounded by the credible interval.

Ultimately, the Bayesian model presented herein provided posterior distributions for each parameter mean. The *median* value of these posterior distributions was selected as the point estimates for the *means* of each parameter. The median value was taken to be a better measure of the central tendency of the posterior distributions in cases where these distributions were asymmetrical.

The empirical data will be presented along with the Bayesian estimates. The empirical data for the participant-level data simply constitute the mean (or log SD in the case of the noise functions) of the dependent variable across trials and the mean of the dependent variable across participants in the case of the group-level data. The purpose of presenting the empirical data along with the estimates is to understand the extent to which the model overcomes (1) the inherent noisiness of the data to generate smooth and accurate estimates, and (2) the adverse impact of outliers in estimating the population-level functions. Also, the presentation of the empirical data allows the reader to examine the extent to which the other methods of analysis, presented in the previous chapter, were adversely impacted by these two factors.

4.2 RESULTS & DISCUSSION

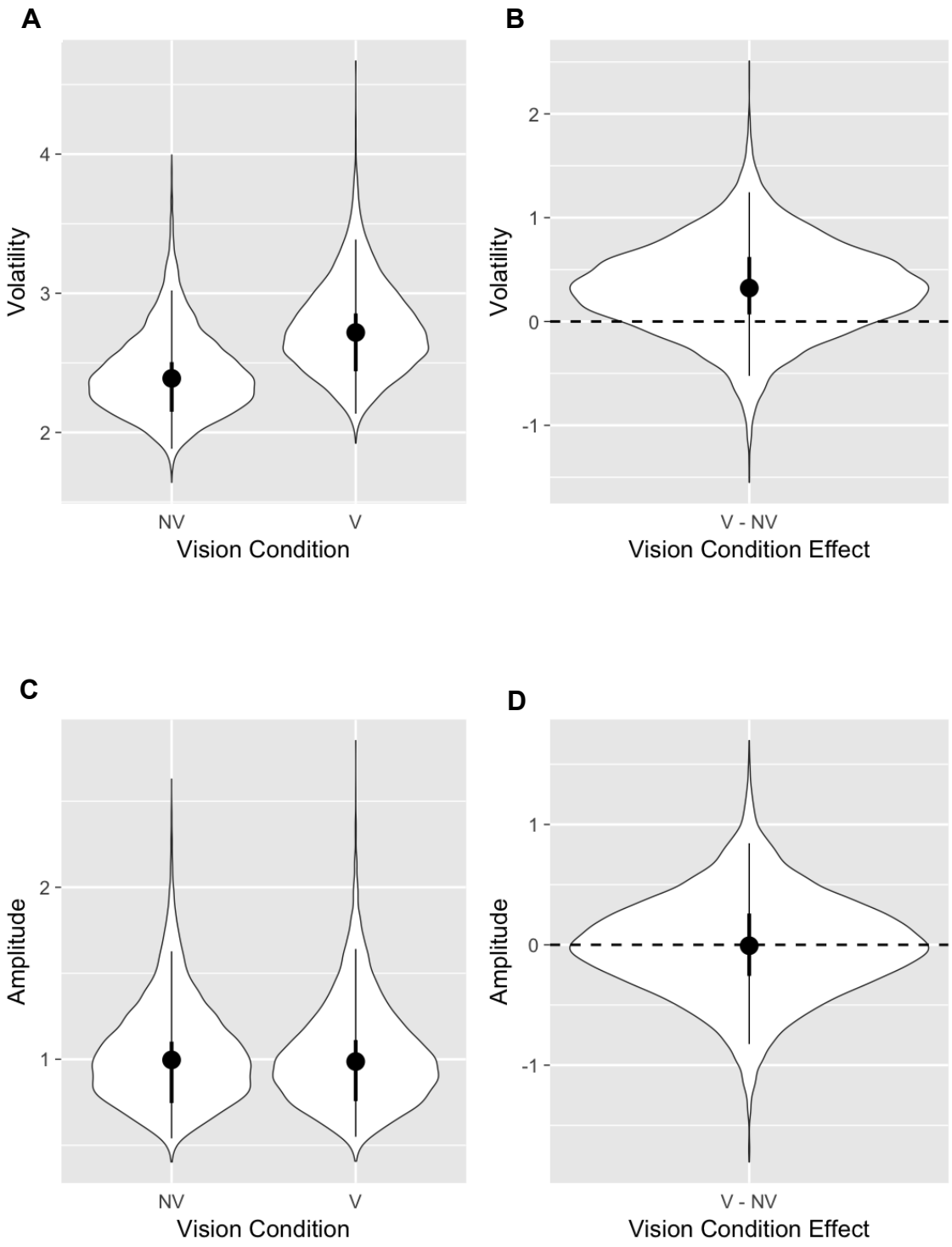
Sixteen chains were run, with 1000 iterations per chain. Half the iterations for each chain were warm-up samples. The analysis took approximately 120 hours to run. The population hyperparameter estimates, the participant-to-participant variability estimates, and the population noise functions were sampled effectively all estimates achieving at least 1000 effective samples and having a corresponding Rhat of 1. The participant mean functions were largely properly sampled with all estimates having a corresponding Rhat of 1. However, some estimates had fewer than 1000 effective samples, the minimum being 400. The participant noise functions were also largely properly sampled with nearly all estimates having a corresponding Rhat of 1. The most deviant Rhat was 1.02. Further, virtually all estimates had at least 1000 effective samples, the minimum being 927. Finally, the population functions were moderately well sampled. The Rhat values ranged from 1.01 to 1.06. Further, the effective samples ranged from 277 to 1137.

4.2.1 Hyperparameters for Mean Functions

The hyperparameter estimates for the population and participant mean functions are displayed in Figure 12. Population volatility was slightly greater in the vision condition than it was in the no-vision condition. The effect of vision condition excluded the 50% credible interval. Although the strength of the effect was weak, the direction of the effect was consistent with the expectation that the population mean function would be more

volatile in the vision condition due to the propensity of this condition to generate a greater number of online corrections – either continuous (impulse control) or discrete (limb-target; Elliott et al, 2017). However, as seen in Figure 16, the population mean functions overall were not particularly volatile (i.e. they were simple and smooth functions). Therefore, the value of examining the hyperparameter values in this context is unclear. Another consideration is whether the downsampling of the trajectories prior to subjecting them to the GPR analysis obfuscates the true volatility effects (for discussion, see subsection 6.5).

There was decidedly no effect of vision condition on population amplitude. This null effect is not surprising given that participants started and ended their movements at the same points in space and generally did not overshoot the target because of the simplicity of the task. Population amplitude variability was slightly greater in the no-vision condition than it was in the vision condition. The effect of vision excluded the 50% credible interval. Although the effect was small, it may be explained in the context of the effect of vision condition on VE (amplitude). Finally, there was no effect of vision condition on population volatility variability. No particular prediction had been made with respect to this hyperparameter measure.


A
Volatility
4
3
2
NV
V
Vision Condition
B
Volatility
2
1
0
-1
V - NV
Vision Condition Effect
C
Amplitude
2
1
NV
V
Vision Condition
D
Amplitude
1
0
-1
V - NV
Vision Condition Effect

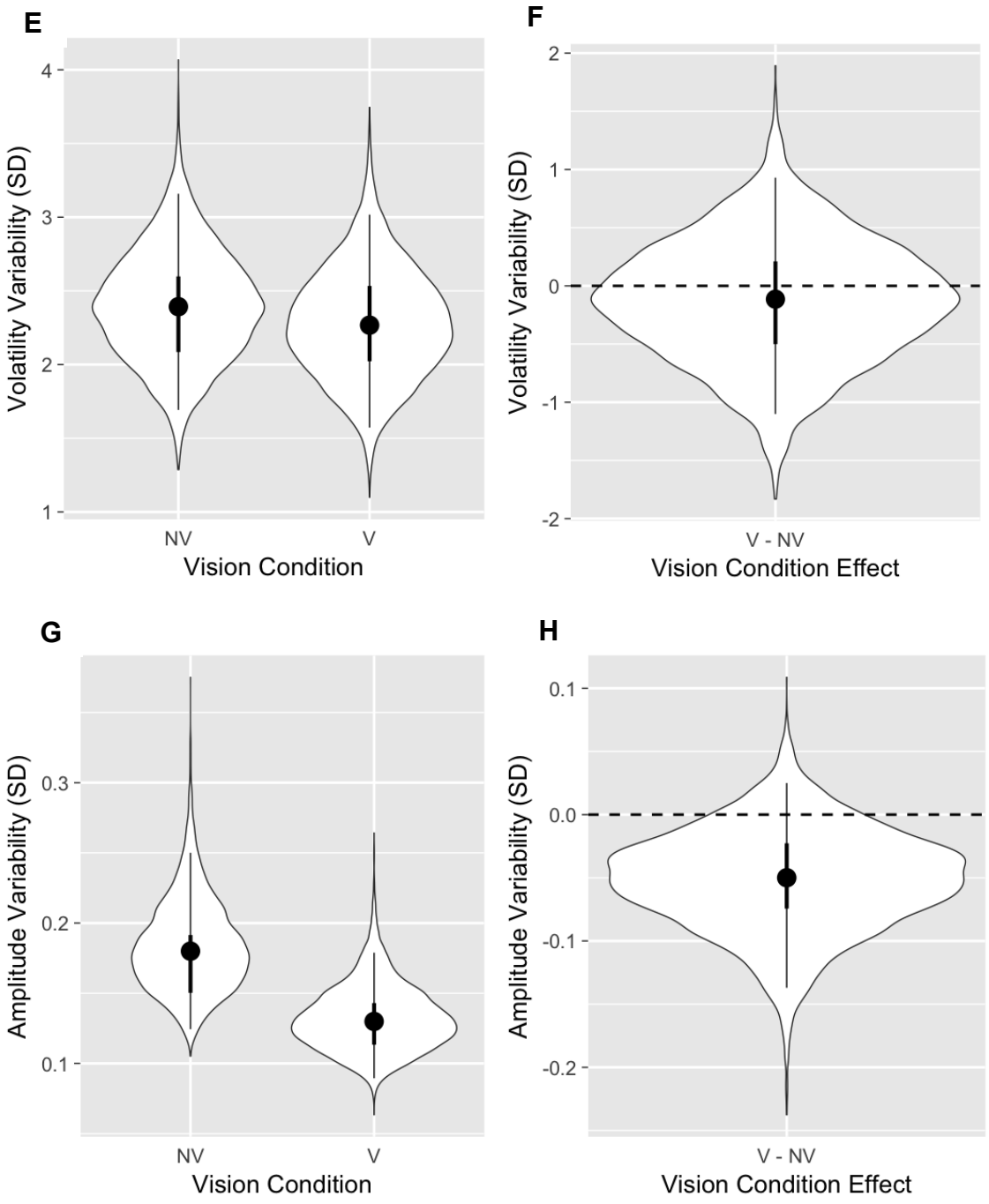
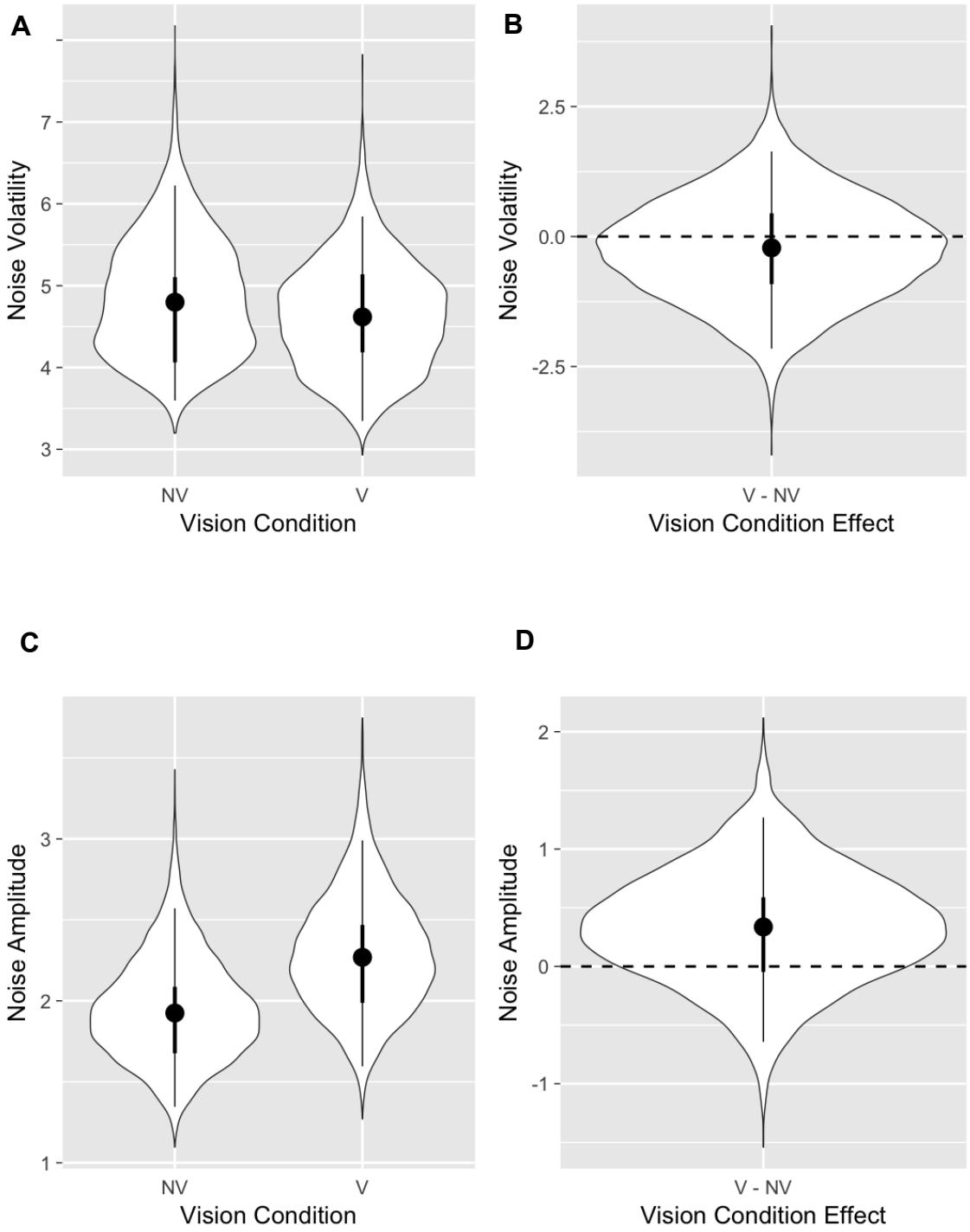

E
Volatility Variability (SD)
4
3
2
1
NV
V
Vision Condition
F
Volatility Variability (SD)
2
1
0
-1
-2
V - NV
Vision Condition Effect
G
Amplitude Variability (SD)
0.3
0.2
0.1
NV
V
Vision Condition
H
Amplitude Variability (SD)
0.1
0.0
-0.1
-0.2
V - NV
Vision Condition Effect

Figure 12 Violin plots of population volatility (A & B) and amplitude (C & D) estimates, and of participant volatility (E & F) and amplitude (G & H) variability estimates. Posterior distributions are mirrored along the vertical axis for each estimate. For each row, the left graph displays the estimates as a function of vision condition whereas the right graph displays the effect of vision condition (vision minus no-vision). The dotted black line on the right graph represents a null effect (zero). The large black dot of each estimate represents the median of the posterior distribution. The thick black lines represent 50% credible intervals and the thin black lines represent 95% credible intervals. The no-vision and vision conditions are represented by "NV" and "V", respectively.

4.2.1 Hyperparameters for Noise Functions

The hyperparameter estimates for the population and participant noise functions are displayed in Figure 13. There were no effects of vision condition on population noise volatility and noise amplitude. There were also no effects of vision on participant noise volatility and amplitude variability. One could have expected an effect of vision condition on population noise amplitude because trial-wise movement variability was greater in the vision condition in the early parts of the movement (see Figures 8, 9, and 19).


A
Noise Volatility
7
6
5
4
3
NV
V
Vision Condition
B
Noise Volatility
2.5
0.0
-2.5
V - NV
Vision Condition Effect
C
Noise Amplitude
3
2
1
NV
V
Vision Condition
D
Noise Amplitude
2
1
0
-1
V - NV
Vision Condition Effect

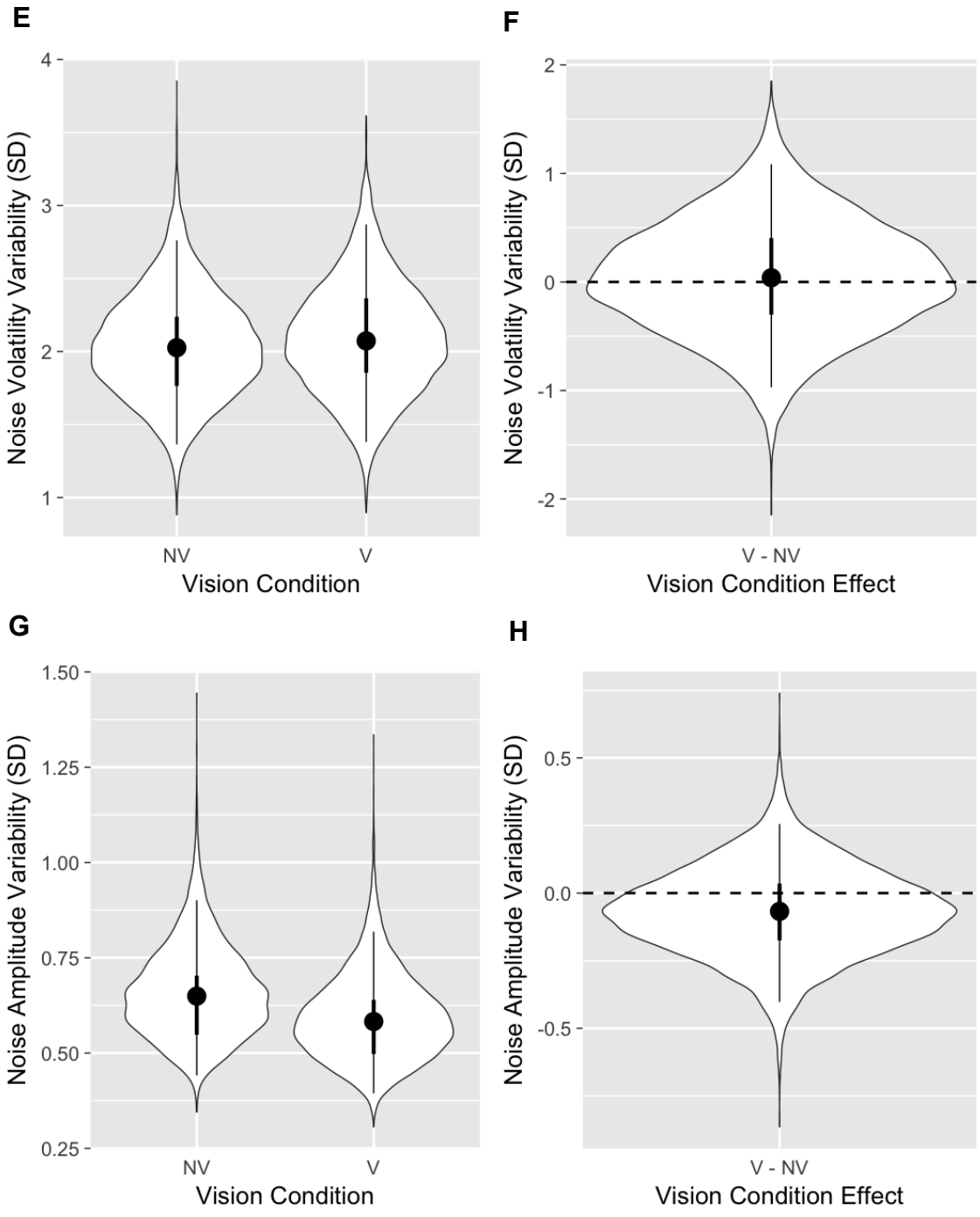
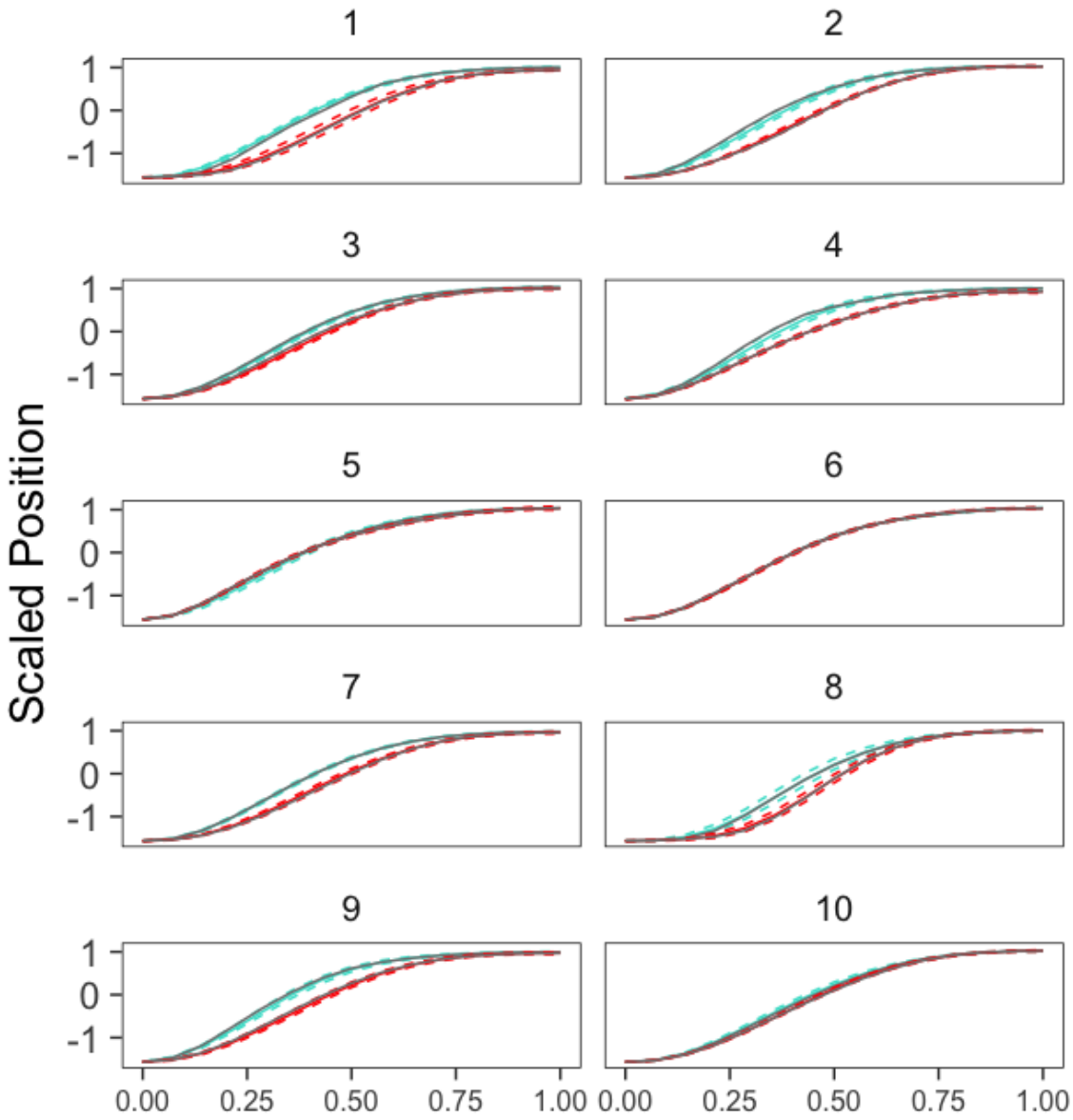
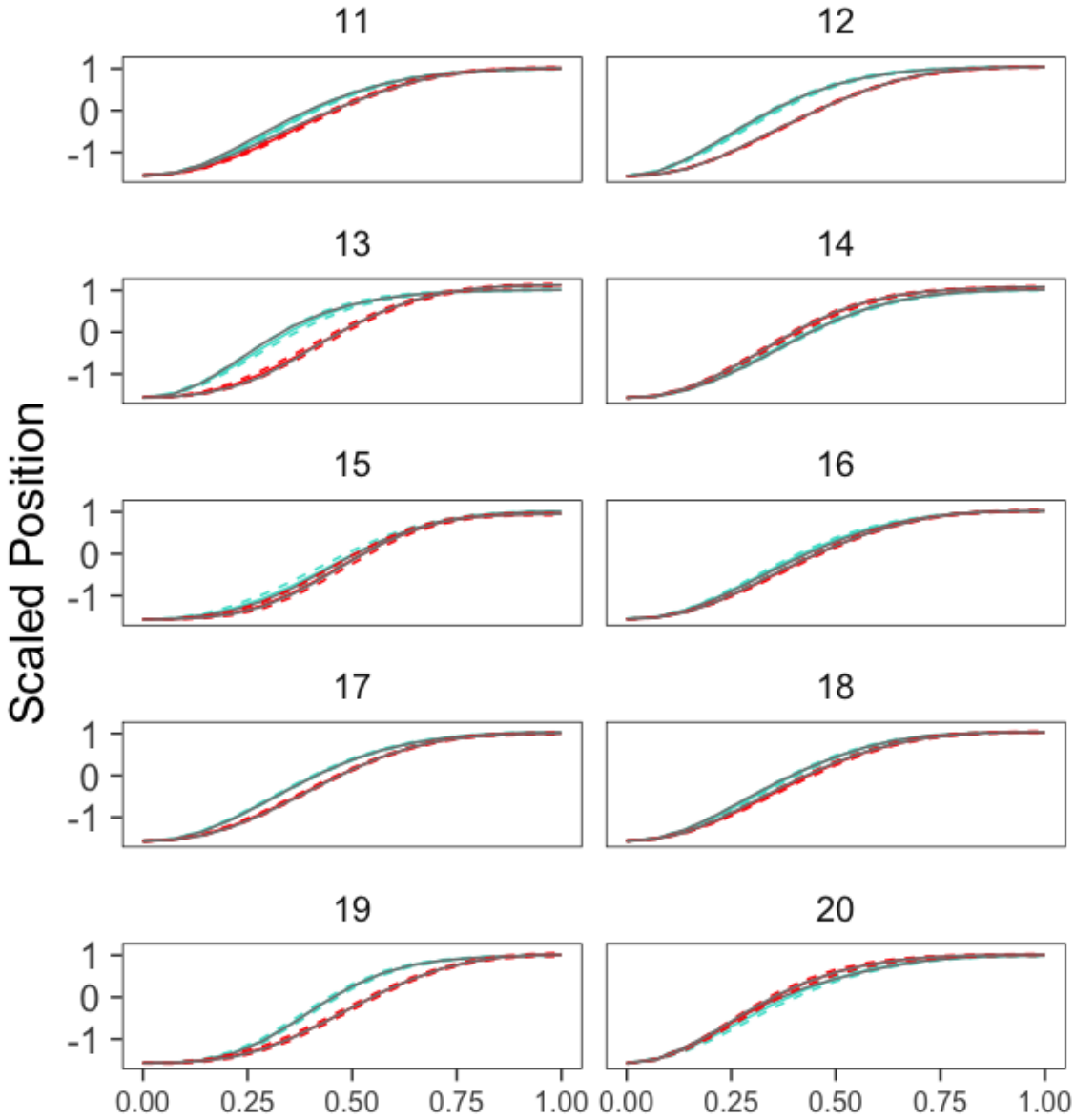

E
Noise Volatility Variability (SD)
4
3
2
1
NV
V
Vision Condition
F
Noise Volatility Variability (SD)
2
1
0
-1
-2
V - NV
Vision Condition Effect
G
Noise Amplitude Variability (SD)
1.50
1.25
1.00
0.75
0.50
0.25
NV
V
Vision Condition
H
Noise Amplitude Variability (SD)
0.5
0.0
-0.5
V - NV
Vision Condition Effect

Figure 13 Violin plots of population noise volatility (A & B) and noise amplitude (C & D) estimates, and of participant noise volatility (E & F) and noise amplitude (G & H) variability estimates. Posterior distributions are mirrored along the vertical axis for each estimate. For each row, the left graph displays the estimates as a function of vision condition whereas the right graph displays the effect of vision condition (vision minus no-vision). The dashed black lines on the graphs on the right represents a null effect (zero). The large black dot of each estimate represents the median of the posterior distribution. The thick black lines represent 50% credible intervals and the thin black lines represent 95% credible intervals. The no-vision and vision conditions are represented by "NV" and "V", respectively.

4.2.3 Participant Mean Functions

The participant mean trajectory estimates are presented in Figure 14. The participant-wise mean trajectory estimates captured the empirical means of the normalized participant-wise trajectories. This observation indicates, albeit in a crude manner, that the participant-wise estimates represent the raw data reasonably well. Figure 15 displays the median values of the posterior estimates across participants for each condition to facilitate the detection of outliers. There appears to be an outlying participant in the no-vision condition (pink trajectory). This participant covers most of the distance to the target in the early portion of their movement. Upon reviewing the trial-by-trial data, it appears as though the outlying trajectory is a result of the outlying participant overshooting the target substantially more often than the other participants.


1
2
3
4
5
6
7
8
9
10
Scaled Position
1
0
-1
0.00
0.25
0.50
0.75
1.00


11
12
13
14
15
16
17
18
19
20
Scaled Position
1
0
-1
0.00
0.25
0.50
0.75
1.00

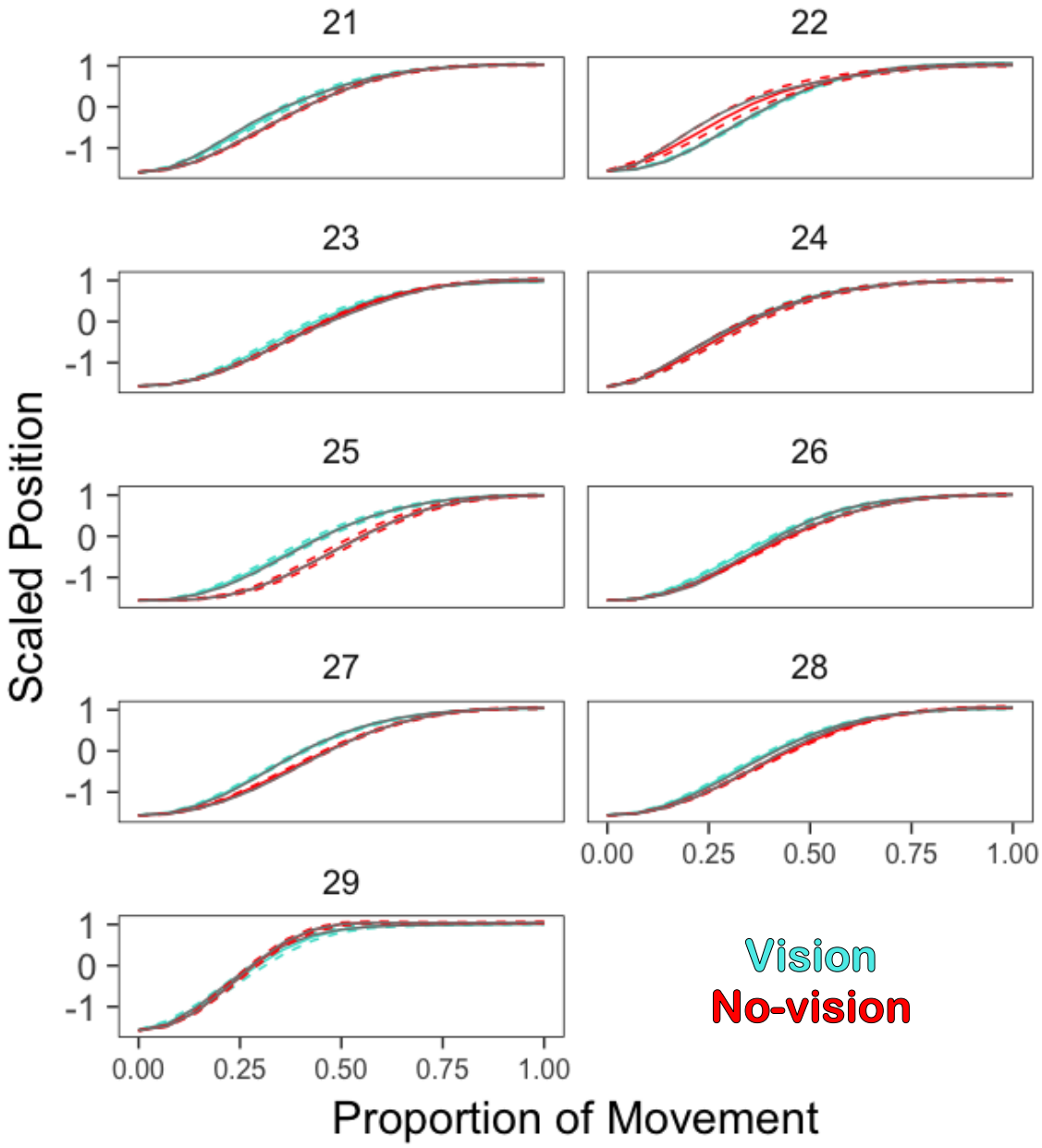


Figure 14 The mean function estimates for each participant, for vision (turquoise) and no-vision (red) conditions, with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant mean functions. The dashed colored lines, surrounding the solid colored lines, make up the 95% credible intervals of the participant mean function estimates. The solid light gray lines represent the empirical means of the normalized data.

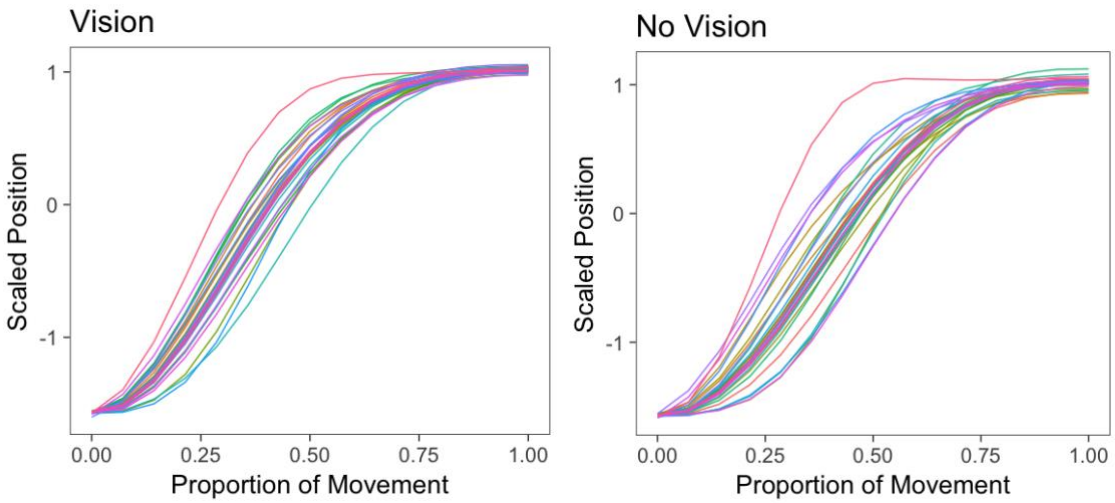
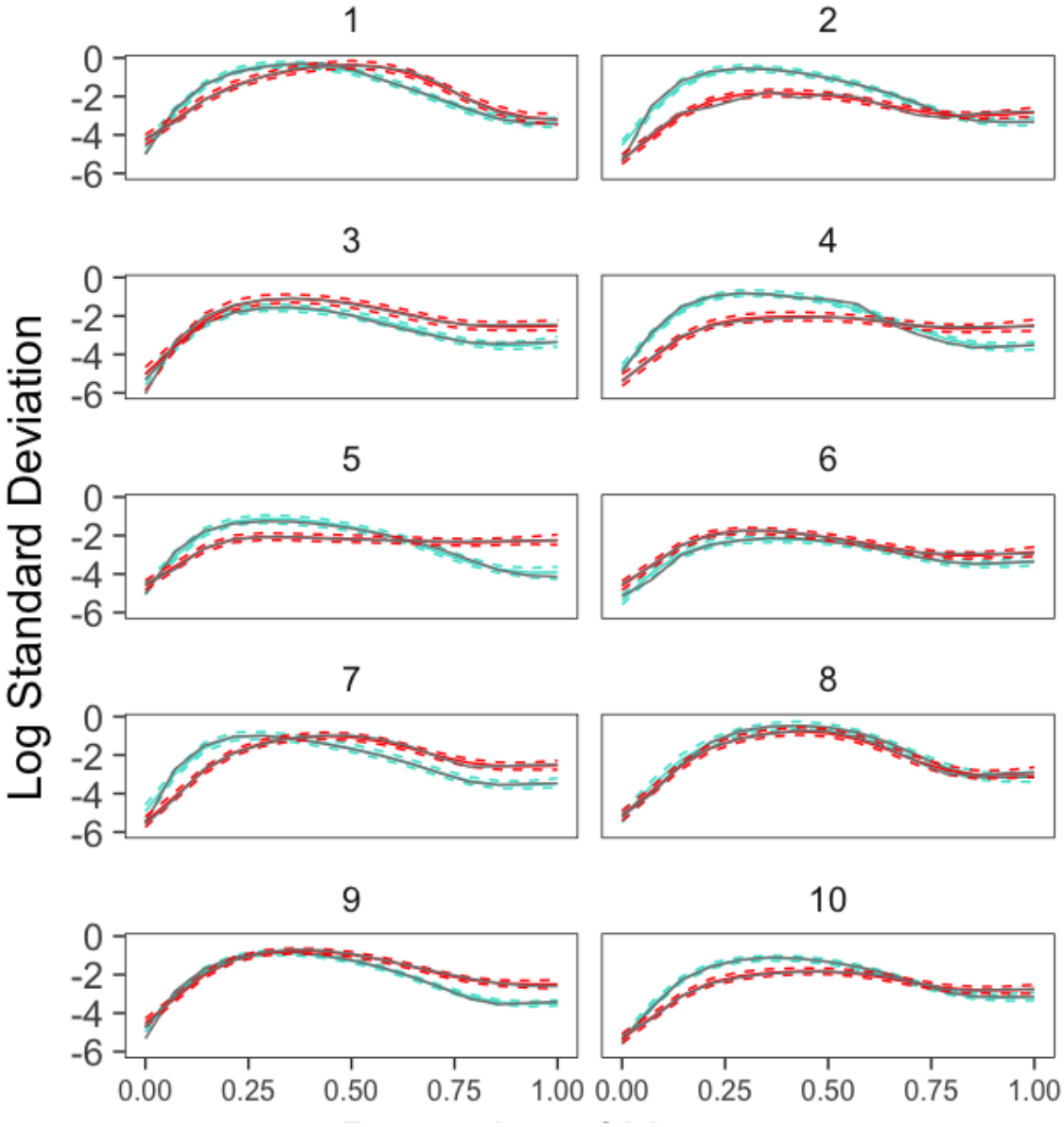


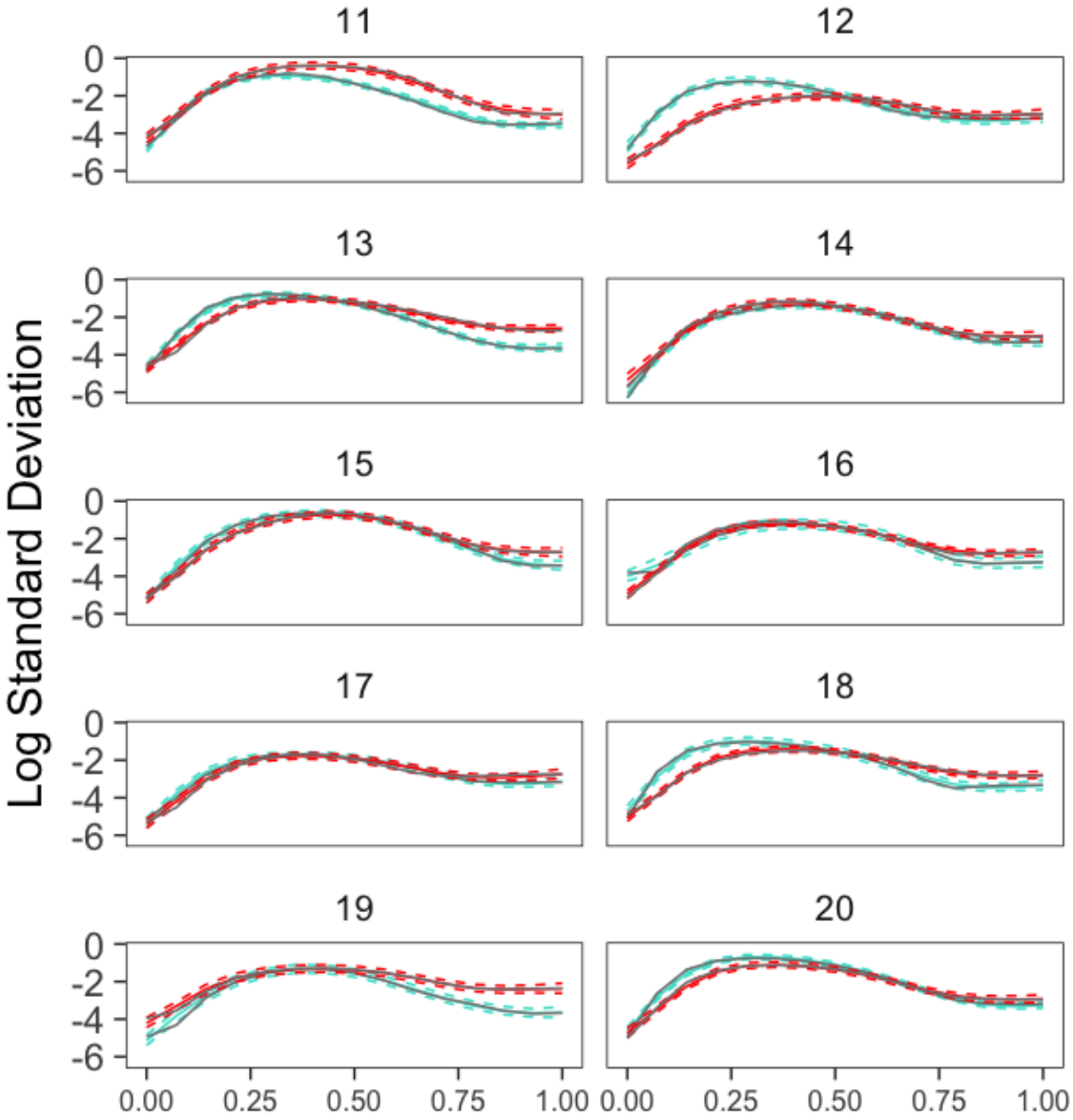
Figure 15 The mean function estimates for each participant, for the vision and no-vision conditions, with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant mean functions.

4.2.4 Participant Noise Functions

The participant noise trajectory estimates are presented in Figure 16. The participant-wise noise function estimates captured the empirical noise functions of the normalized participant-wise trajectories. As with the participant mean trajectory estimates, the match between the empirical and estimated noise functions indicates that the model is largely basing its estimates on the data and not being overly influenced by priors and a complex model structure. Figure 17 displays the median values of the posterior estimates across participants for each condition to facilitate the detection of outliers. Unlike with the

participant mean trajectory estimates, there were no notably deviant participant-wise noise functions.




11
12
13
14
15
16
17
18
19
20
Log Standard Deviation
0
-2
-4
-6
0.00
0.25
0.50
0.75
1.00

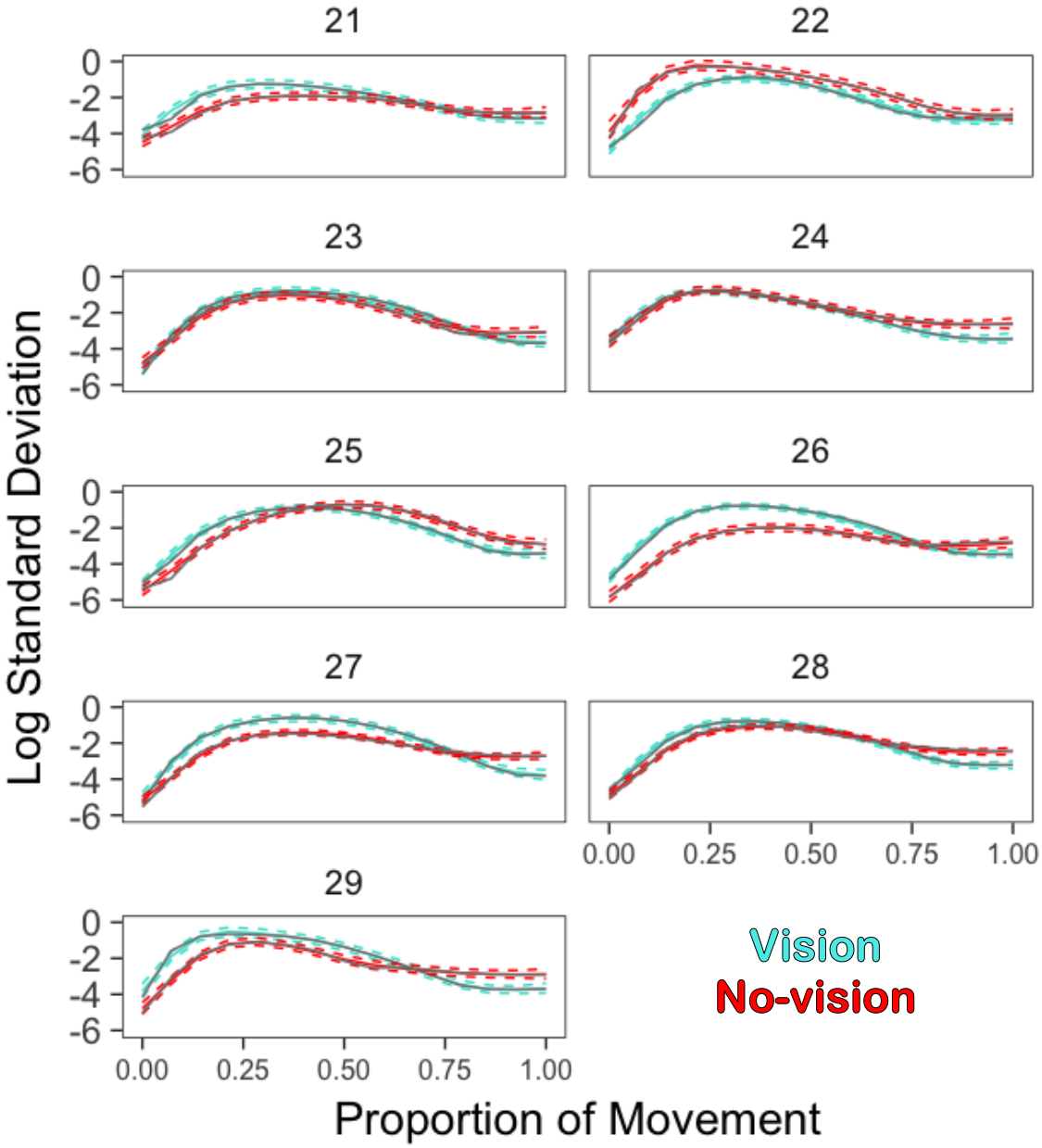


Figure 16 The mean noise function estimates for each participant, for vision (turquoise) and no-vision (red) conditions, with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant noise functions. The dashed colored lines, surrounding the solid colored lines, make up the 95% credible intervals of the participant noise function estimates. The solid light gray lines represent the empirical means of the normalized data.

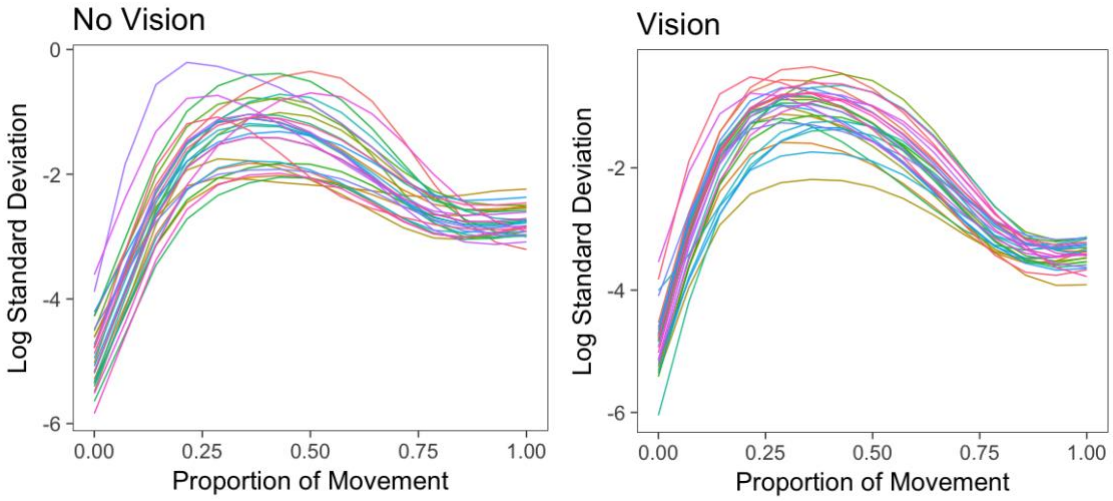


Figure 17 The noise function estimates for each participant, for the vision and no-vision conditions, with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant noise functions.

4.2.5 Population Mean Functions

The population mean trajectory estimates are presented in Figure 18. There was large effect of vision condition on the population mean functions such that participants covered the distance required to reach the target earlier in their proportional movements in the vision condition than they did in the no-vision condition. The effect was greatest around the time of PV. This result is consistent with what was observed with the fANOVA analysis (see Figure 10). Indeed, it is difficult to compare the estimates of the fANOVA analysis method with those of the Bayesian method because both sets of estimates contained little uncertainty (i.e. both analysis methods, in this case, produced precise

estimates of the population mean trajectories). Further, the error bars for each analysis are not directly comparable to each other as the fANOVA error bars were FLSDs and thus appropriate for detection of significant differences between conditions, whereas the GPR error bars simply indicate the credible range of values for each condition which is why a separate effect function is plotted for the explicit determination of the effects of vision. Bayesian estimates largely captured the empirical means of the normalized trajectories thereby providing a partial confirmation that the raw data are appropriately represented by the model.

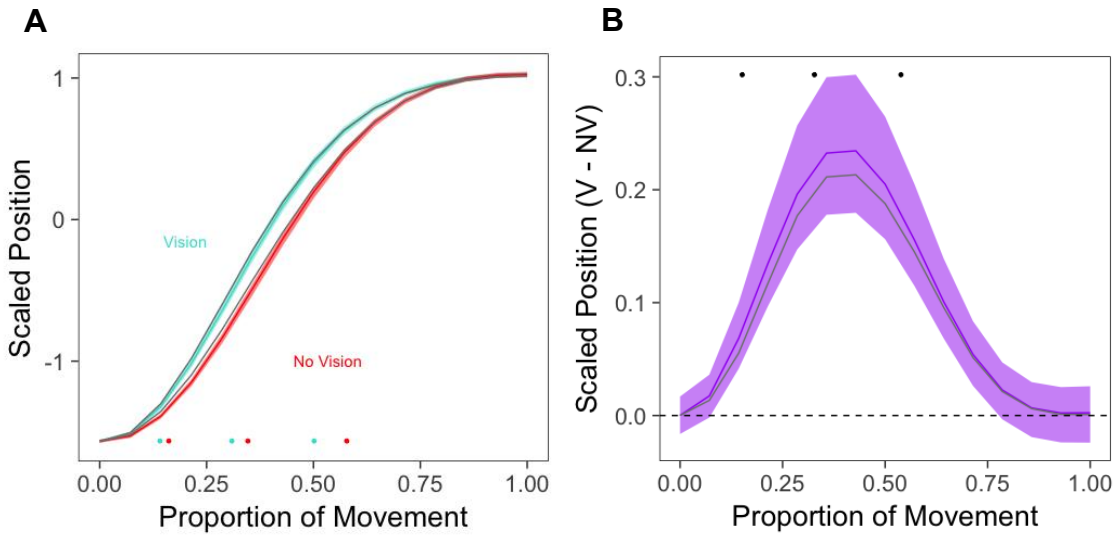


Figure 18 The population mean function estimates for the vision and no-vision conditions (A), and the effect of vision condition (vision minus no-vision; B), with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates. The error bands surrounding the solid colored lines make up the 95% credible intervals of the posterior estimates. The solid light gray lines represent the empirical means of the normalized data. The dots along the bottom (A) or top (B) of the graphs correspond to the mean kinematic marker estimates (from left to right: PA, PV, PD). The no-vision and vision conditions are represented by "NV" and "V", respectively.

4.2.6 Population Noise Functions

The population noise trajectory estimates are presented in Figure 19. There was large effect of vision condition on the population noise functions. Trial-by-trial variability was greater in the vision condition in the early portion of the movement and greater in the no-vision condition in the later portion of the movement. The effect in the early portion of the movement was greatest around the time of PA.

These results are consistent with those observed with the spatial variability profile and epsilloid methods. Of course, one difference is that the population noise functions from the Bayesian model are estimated in log space. One advantage of this representation is that the uncertainty associated with the estimates is more consistent throughout the movement, facilitating the visual interpretation of effect sizes. In addition, the Bayesian estimates appear to have been more sensitive in detecting spatial variability at towards the end of the movement, compared to the spatial variability profile and epsilloid methods. The Bayesian estimates largely captured the empirical variability of the normalized trajectories. However, the estimate in the vision condition had greater trial-wise variability values early in the movement relative to the empirical data. This discrepancy may have been a result of the mildly outlying participant noise function that has less trial-wise variability than the rest of the functions (see Figure 17). Bayesian hierarchical model estimates tend to be less-so pulled by outlying data than empirical means and

traditional model estimates (for discussion, see Kruchke, 2014).

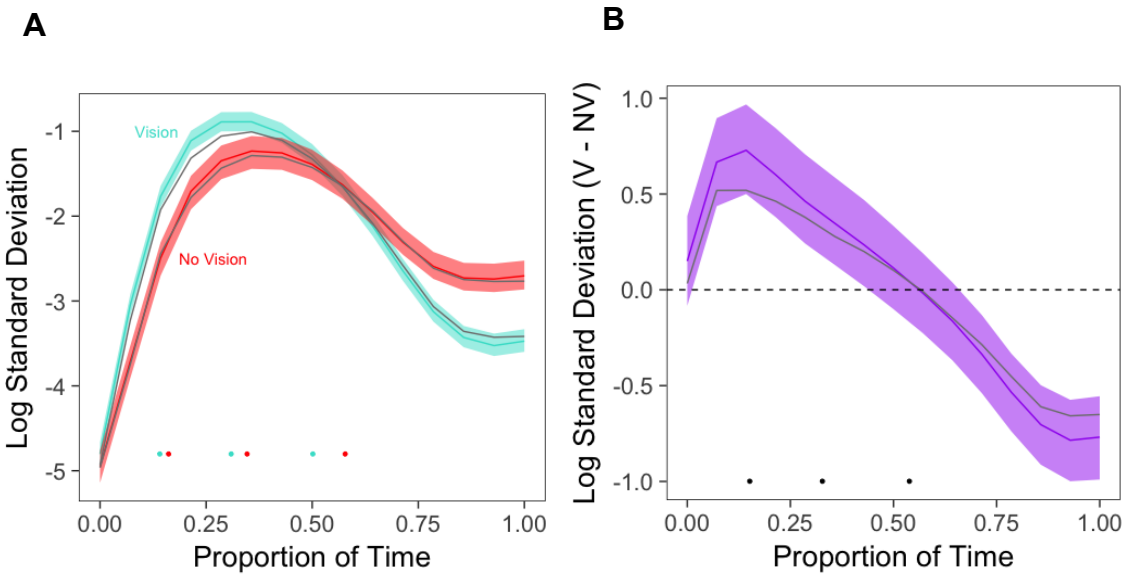


Figure 19 The population noise function estimates for the vision and no-vision conditions (A) and the effect of vision condition (vision minus no-vision; B), with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates. The error bands surrounding the solid colored lines make up the 95% credible intervals of the posterior estimates. The solid light gray lines represent the empirical variability (log SDs) of the normalized data. The dots along the bottom (A) or top (B) of the graphs correspond to the mean kinematic marker estimates (from left to right: PA, PV, PD). The no-vision and vision conditions are represented by "NV" and "V", respectively.

4.3 CONCLUSION

The hyperparameter estimates did not indicate any substantial effects of vision. Perhaps the one perhaps noteworthy hyperparameter result was a moderately credible effect of vision on population mean function volatility such that the function in the vision

condition was more volatile than the one in the no-vision condition.

The participant mean and noise function estimates tracked the empirical data closely. Therefore, the model appeared to represent to participant-level data accurately. The population mean function estimates tracked the empirical data well and indicated a large effect of vision that was maximized around the time of PV: the participants covered the distance required to reach the target earlier in their proportional movements in the vision condition than they did in the no-vision condition. This result was consistent with the result of the fANOVA analysis of Chapter 3. Both analysis methods generated precise estimates, making difficult to identify an advantage of one over the other.

The population noise function estimates tracked to empirical data reasonably well and demonstrated that participant trajectories had greater trial-wise variability in the vision condition in the early phase of the movement, and less trial-wise variability in the later phase of the movement. This result was consistent those of the spatial variability and epsilloid analyses of Chapter 3. However, the GPR estimates were more precise and nuanced, and therefore detected effects of vision that were not identified by the other two methods.

CHAPTER 5 SIMULATION

5.1 METHODS

Before presenting the results of the simulated dataset, it is worth clarifying the purpose of the simulation. First, the simulation provided a method of validating the Bayesian hierarchical GPR model. The Bayesian model is complex and dependent on MCMC sampling methods for its implementation. Its complexity makes researchers prone to model misspecifications, whereas the lack of an analytical solution entails the potential for bad approximations of the posterior distributions. A large-scale simulation, whereby data are generated from a known generative model and then estimated by an analytical model of interest, allows the researcher to validate that the analytical model can properly estimate the true parameter values from the sample data. The second purpose of the simulation was to compare the ability of different statistical methods to estimate the true underlying effects of a population or a subject from the sample data. In an experimental setting, it is difficult to compare the validity of statistical methods since one does not have access to the value of the real effects being estimated. In a simulation setting, one has access to and ultimately determines that information.

The simulated data was sampled from the generative model that was most theoretically plausible. It had the same form as the analytical model (for a detailed description of the model, see subsection 4.1). The simulated dataset resembled the experimental dataset in its structure: 29 participants, each split into two conditions with 20 trials per condition.

The simulated trajectory data from either condition were sets of arbitrary functions. The generative model allowed for participant-wise deviations in these functions. It also allowed for participant-wise deviations in the variability of the measurement across trials. Finally, the variability across trials itself will be allowed to vary along with the time course of the trajectory. Each trial consisted of a trajectory was made up of 15 samples. The population parameters of the generative model were selected so that the trajectories at least broadly resembled typical movement trajectories: fairly smooth and not too volatile.

The population mean and trial-wise variability functions of the two hypothetical experimental conditions were sampled from GPs with set hyperparameters. Most of these hyperparameters were set to differ slightly. The other model parameters were also set to differ mildly between the two conditions. The true parameter values will be presented graphically along with their model estimates subsection 5.2. This will allow the reader to visualize the effects that were built into the generative model. Data were only simulated in one dimension because computational limitations precluded the modelling of multiple dimensions simultaneously. As such, the simulated trajectories were one-dimensional.

Importantly, the two conditions were not meant to represent vision and no-vision conditions from a standard goal-directed reaching experiment. As stated, the shape of the simulated trajectories were arbitrary. Simulating functions from a GP prior involves setting the hyperparameter values (volatility, amplitude, and the extent to which these vary across participants), then randomly sampling functions from the corresponding GP. The randomness involved in sampling these functions makes it difficult to readily

simulate trajectories with specific characteristics. For the purposes of the present study, datasets were simulated until one in which there were reasonable differences between conditions and reasonable levels of trial-wise variability was identified.

The data analysis methods for the simulated dataset were virtually identical to those applied to the experimental dataset. The primary difference between the analyses of the experimental and simulated datasets was that fewer analysis methods were applied to the latter. In particular, the outcome variables – RT, MT, VE, CE, and target error – were omitted from the following analysis because the simulation did not generate these DVs. Further, the TAPV and discontinuity analyses were omitted because the former was not explicitly manipulated in generating the simulated trajectories, whereas the latter was obsolete since the simulated trajectories had no discontinuities. Another difference between the analyses of the experimental and simulated datasets was that, for the simulated data, it was not meaningful to extract kinematic markers such as PA, PV, and PD for the R^2 and spatial variability profile analyses. Instead, the position at END was predicted from the positions at each time point. Finally, for the fANOVA analysis, since the simulated data were made up of trajectories with 15 samples, the position values for movement amplitude were sampled from 15 (instead of 200) normalized time points, equally-spaced within each trial. The results of these analyses, the fANOVA analysis, and the Bayesian hierarchical GPR analysis are presented in the following subsection.

5.2 RESULTS & DISCUSSION

5.2.1 R^2

The R^2 estimates as a function of condition and proportion of movement are presented in Figure 20. The R^2 analysis did not reveal any significant effects. Therefore, the R^2 analysis could not detect the population volatility effect that was instantiated in the generative model. This may have been a result of the R^2 values being low across all time points. Alternatively, this may have been due to the fact that the R^2 analysis does not measure the same thing as the volatility hyperparameter.

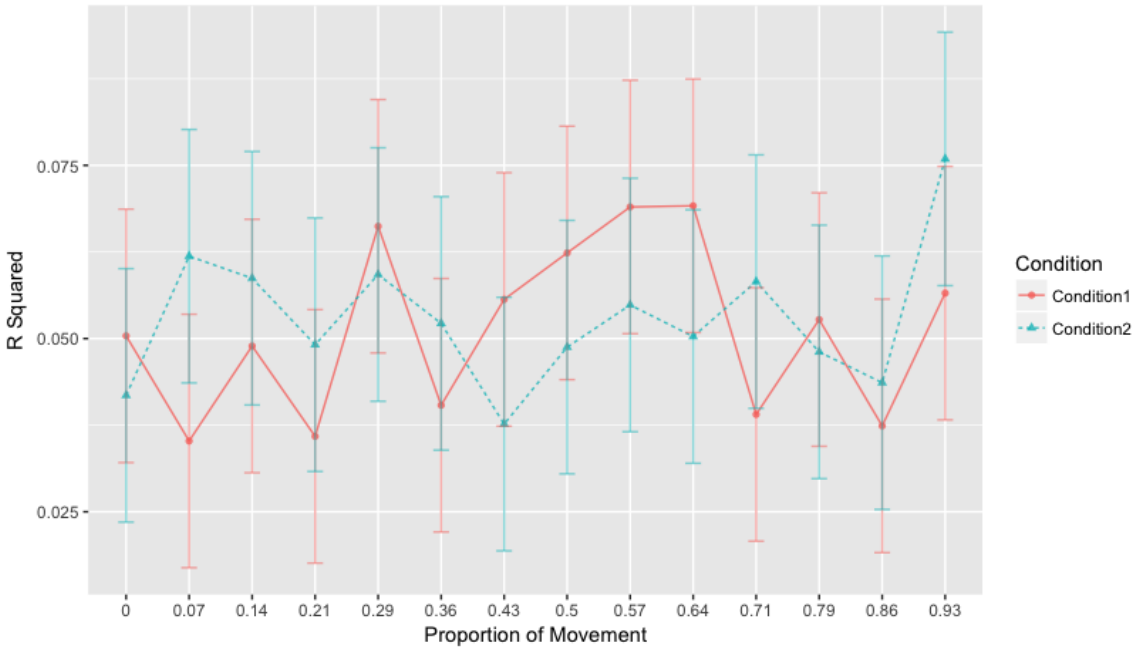


Figure 20 R^2 values plotted as a function of proportion of movement (x-axis) and condition (line/point color and type). Error bars represent FLSDs.

5.2.2 Spatial Variability Profile

The spatial variability estimates as a function of condition and proportion of movement are presented in Figure 21. The spatial variability profile estimates captured the true population spatial variability profiles fairly well. The spatial variability profile analysis revealed a main effect of condition [$F(1,28) = 70.3$, $p < 0.001$, $\eta_G^2 = 0.436$] and a main effect of proportion of movement [$F(14,392) = 55.3$, $p < 0.001$, $\eta_G^2 = 0.240$]. The analysis also revealed a condition by proportion of movement interaction effect [$F(14,392) = 80.9$, $p < 0.001$, $\eta_G^2 = 0.375$]. The interaction effect was characterized by greater spatial variability in condition 2 compared to condition 1 later in the movement. However,

Macuhly's test for sphericity was violated for both the main effect of proportion of movement [$W < 0.001$, $p < 0.001$] and the interaction effect [$W < 0.001$, $p < 0.001$]. Using Greenhouse-Geisser corrections, the main effect of proportion of movement [$\varepsilon = 0.202$, $p < 0.001$] and the interaction effect [$\varepsilon = 0.165$, $p < 0.001$] remained.

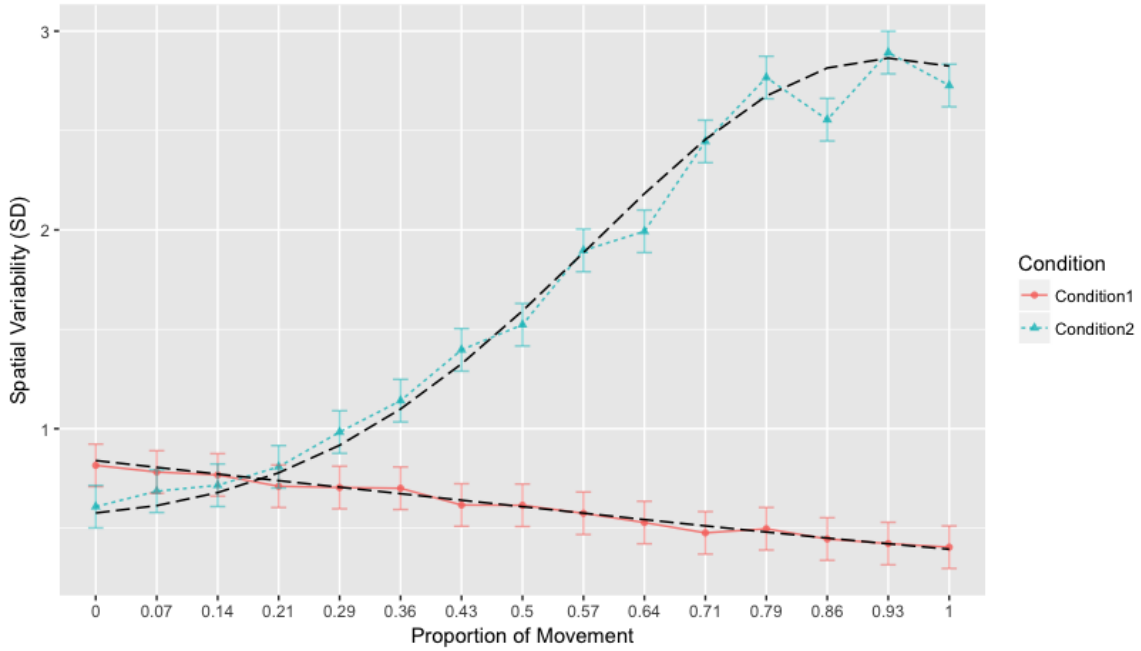


Figure 21 Spatial variability (SD) values plotted as a function of proportion of movement (x-axis) and condition (line/point color and type). Error bars represent FLSDs. The black dashed lines represent the real, latent, population spatial variability profiles.

5.2.3 fANOVA

The fANOVA results are displayed in Figure 22. The fANOVA estimates captured the

true mean population functions fairly well. The condition 2 estimate mildly underestimated the true mean population function. This observation may be explained by the mildly outlying participant presented in Figure 23. The position values of condition 1 were larger than those of condition 2 early in the movement. This difference subsided later in the movement. The spline fitting method appeared to smooth the data – the raw data were notably noisier.

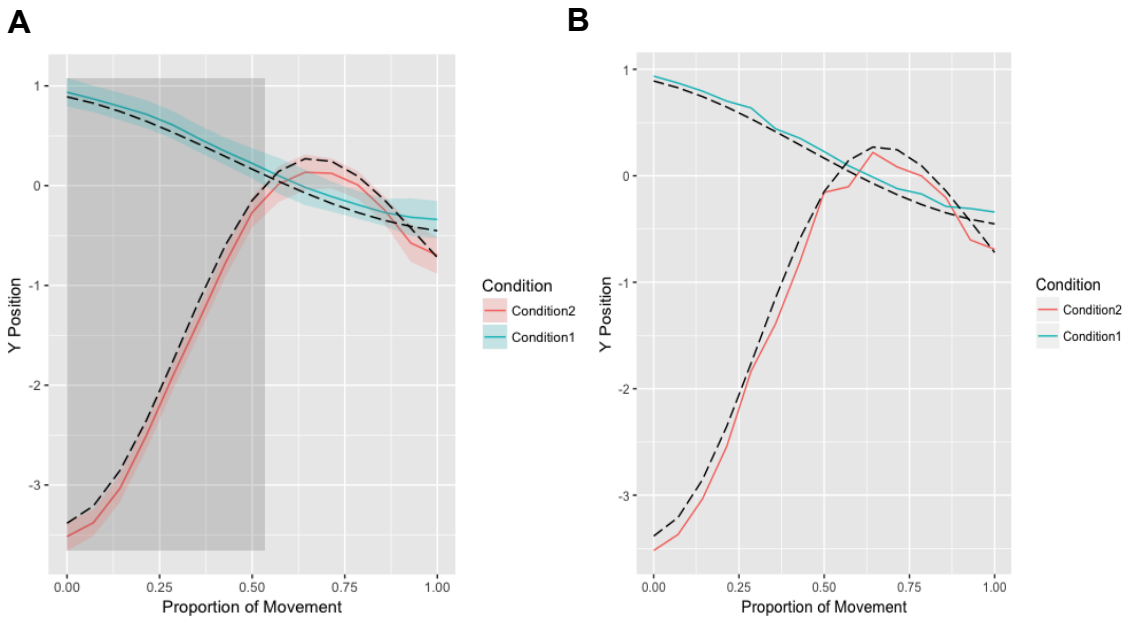


Figure 22 Position values plotted as a function of proportion of movement (x-axis) and condition (line/point color and type) for movement amplitude (y coordinate). Graph A represents the fANOVA. The gray shaded areas represent the time intervals in which the differences between vision conditions were statistically significant. Error bands represent FLSDs, one for each time point. If the error bands failed to overlap at any time point, then the difference at that time point was statistically significant according to an uncorrected t-test. Graph B represents the raw data, which have not undergone smoothing via spline fitting. The black dashed lines represent the real, latent, population mean functions.

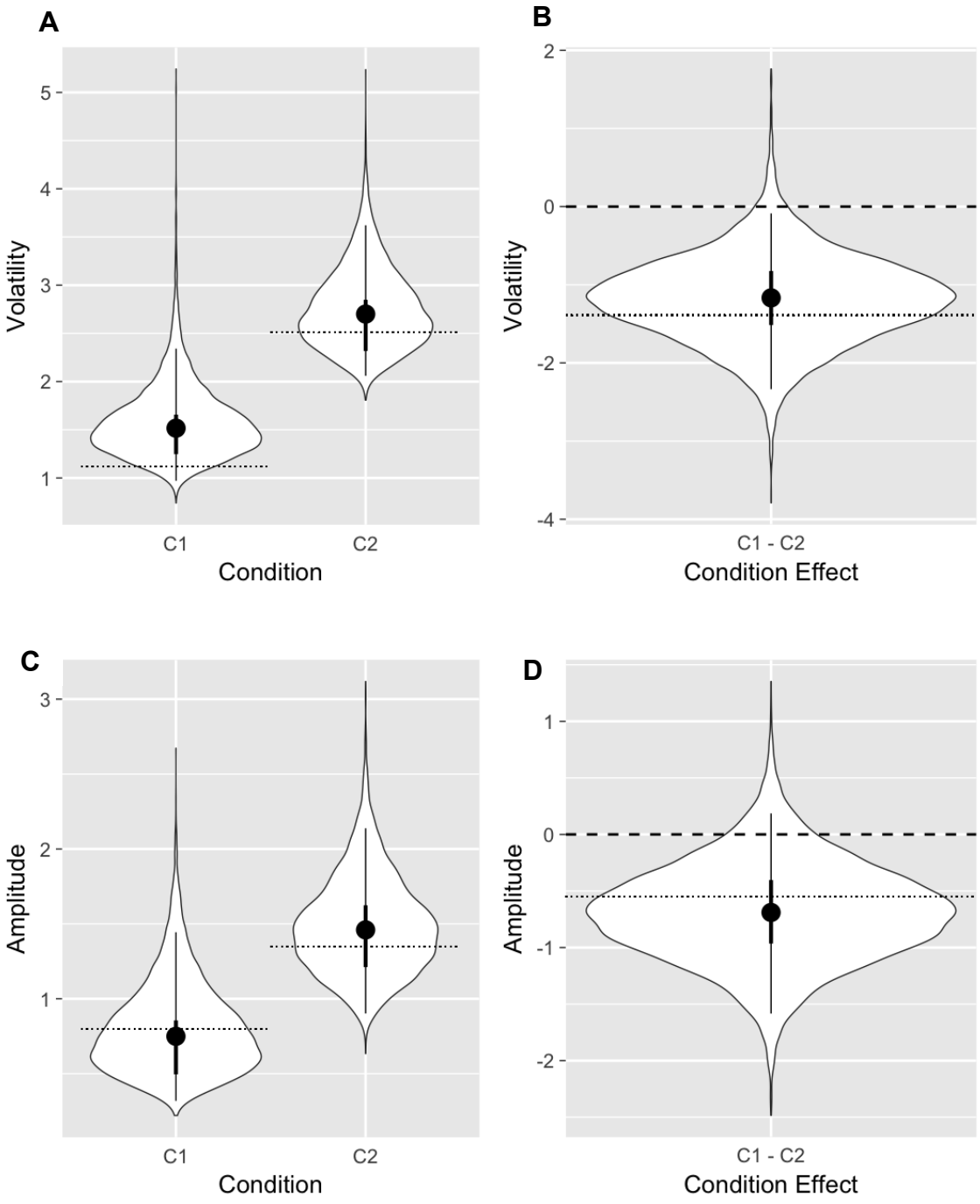
5.2.4 Bayesian Hierarchical Gaussian Process Regression

Sixteen chains were run, with 1000 iterations per chain. Half the iterations for each chain were warm-up samples. The analysis took approximately 10 hours to run. The population hyperparameter estimates, the participant-to-participant variability estimates, and the population mean and noise functions were sampled effectively with all estimates achieving at least 1000 effective samples and having a corresponding Rhat of 1. The participant mean functions were largely properly sampled with all estimates having a corresponding Rhat close to 1. The most deviant estimate had a Rhat of 1.02. Further, some participant mean function estimates had fewer than 1000 effective samples, the minimum being 751. The participant noise functions were also largely properly sampled with nearly all estimates having a corresponding Rhat of 1. The most deviant Rhat was 1.03. Further, virtually all estimates had at least 1000 effective samples, the minimum being 617. 1.01 to 1.06.

5.2.4.1 Hyperparameters for Mean Functions

The hyperparameter estimates of the population and participant mean functions are displayed in Figure 23. To compliment the figure, Table 4 summarizes the extent to which the hyperparameter estimates captured the true population hyperparameter values. The hyperparameter estimates largely captured the true hyperparameter values. However, the population volatility variability 95% credible interval failed to capture the true

hyperparameter value in condition 2. In the current simulation, a reasonable researcher would have properly assessed there to be effects of condition on population amplitude and volatility but failed to identify the existent effects of condition on participant volatility and amplitude variability. Overall, it appeared that the model estimates were more accurate at estimating the population mean hyperparameters than it was at estimating their participant-wise variability (SD). In the author's experience with Bayesian hierarchical modelling, variability (SD) parameters tend to be more difficult to estimate accurately. Fortunately, the latter was not of particular interest in the context of the present experimental research question regarding the effects generated by the vision condition.


A
Volatility
5
4
3
2
1
C1
C2
Condition
B
Volatility
2
0
-2
-4
C1 - C2
Condition Effect
C
Amplitude
3
2
1
C1
C2
Condition
D
Amplitude
1
0
-1
-2
C1 - C2
Condition Effect

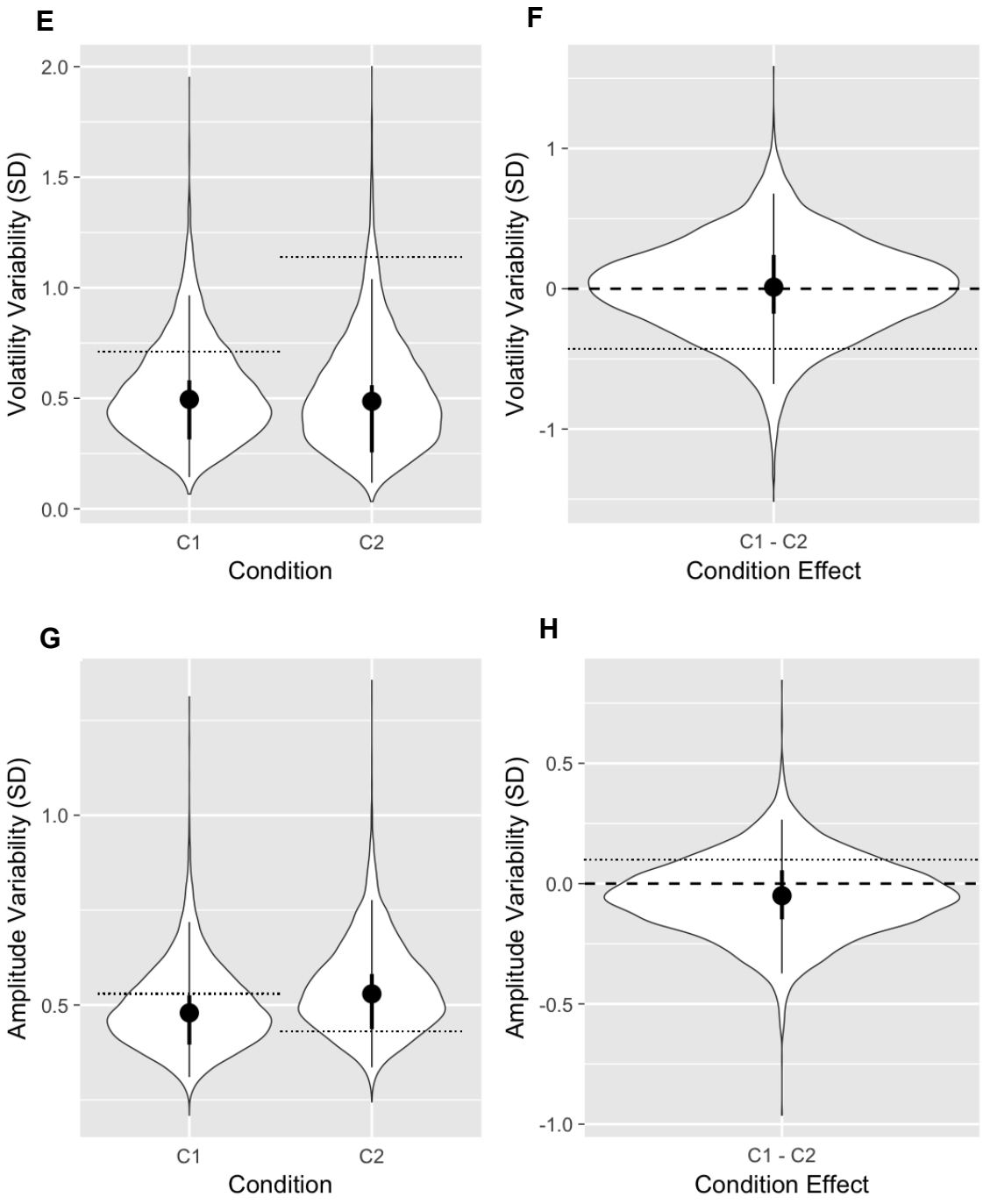

E
Volatility Variability (SD)
2.0
1.5
1.0
0.5
0.0
C1
C2
Condition
F
Volatility Variability (SD)
1
0
-1
C1 - C2
Condition Effect
G
Amplitude Variability (SD)
1.0
0.5
C1
C2
Condition
H
Amplitude Variability (SD)
0.5
0.0
-0.5
-1.0
C1 - C2
Condition Effect

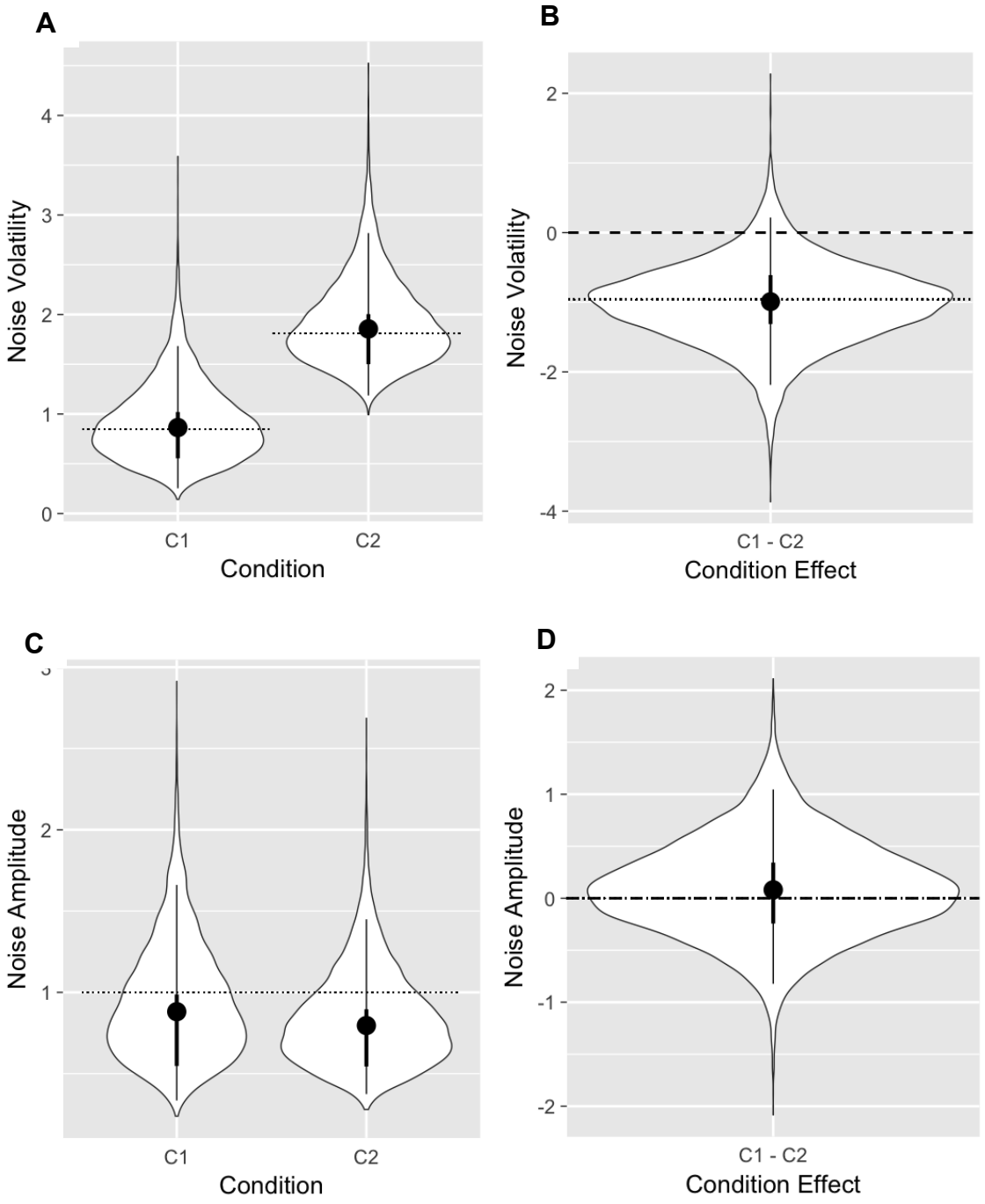
Figure 23 Violin plots of population volatility (A & B) and amplitude (C & D) estimates, and of population volatility (E & F) and amplitude (G & H) variability estimates. Posterior distributions are mirrored along the vertical axis for each estimate. For each row, the left graph displays the estimates as a function of condition whereas the right graph displays the effect of condition (condition 1 minus condition 2). The dashed black lines on the graphs on the right represents a null effect (zero). The horizontal dotted lines are the real hyperparameter values that were used to generate the simulated data. The large black dot of each estimate represents the median of the posterior distribution. The thick black lines represent 50% credible intervals and the thin black lines represent 95% credible intervals. Condition 1 and condition 2 are represented by "C1" and "C2", respectively.

	Volatility	Amplitude	Volatility (SD)	Amplitude (SD)
C1	95%	50%	95%	95%
C2	50%	50%	Excluded	95%
Effect	50%	50%	95%	95%
Zero	Excluded	95%	50%	50%

Table 4 Summary of degree to which population hyperparameter estimates – volatility, amplitude, population volatility variability, and population amplitude variability - for condition 1 (C1), condition 2 (C2), and the effect (C1 minus C2) captured the true population hyperparameter values. If the true hyperparameter value was captured by the 50% credible interval, "50%" was entered into the corresponding box and the box was highlighted as green to convey that the population hyperparameter had been accurately estimated. If the true hyperparameter value was captured by only the 95% credible interval, "95%" was entered into the corresponding box and the box was highlighted as *light* green to convey that the population hyperparameter had been somewhat accurately estimated. Finally, if the true hyperparameter value was not captured by the 95% credible interval, "Excluded" was entered into the corresponding box and the box was highlighted as red to convey that the population hyperparameter had not been accurately estimated. Further, the bottom row of the table was added to indicate the degree to which the estimate of the effect of condition excluded zero (a null effect) in order to gauge the sort of conclusion that would have been drawn regarding the probability of the effect having had no knowledge of the true population effect (i.e. in an experimental setting). The coloring and entry scheme was reversed for this row of the table since the exclusion of zero from the estimate would have corresponded to an accurate assessment of the effect being present with a high degree of probability. Conversely, the inclusion of zero by the 50% credible interval would have corresponded to the erroneous assessment of there probably *not* being an effect when an effect was indeed present.

5.2.4.2 Hyperparameters for Noise Functions

The hyperparameter estimates of the population and participant noise functions are displayed in Figure 24. To compliment the figure, Table 5 summarizes the extent to which the hyperparameter estimates captured the true population hyperparameter values. The hyperparameter estimates largely captured the true hyperparameter values. However, the population noise volatility variability 95% credible interval failed to capture the true hyperparameter value in condition 1. In the current simulation, a reasonable researcher would have properly assessed there to be an effect of condition on population noise volatility and participant noise amplitude volatility, and no effect of condition on population noise amplitude. However, this hypothetical researcher would have failed to identify the existent effect of condition on population noise volatility variability. Overall, it appeared that the model estimates were more accurate at estimating the population mean hyperparameters than it was at estimating their participant-wise variability (SD). Fortunately, the latter is not of particular interest in the context of the present experimental research question regarding the effects generated by vision condition.


A
Noise Volatility
0
1
2
3
4
C1
C2
Condition
B
Noise Volatility
-4
-2
0
2
C1 - C2
Condition Effect
C
Noise Amplitude
1
2
C1
C2
Condition
D
Noise Amplitude
-2
-1
0
1
2
C1 - C2
Condition Effect

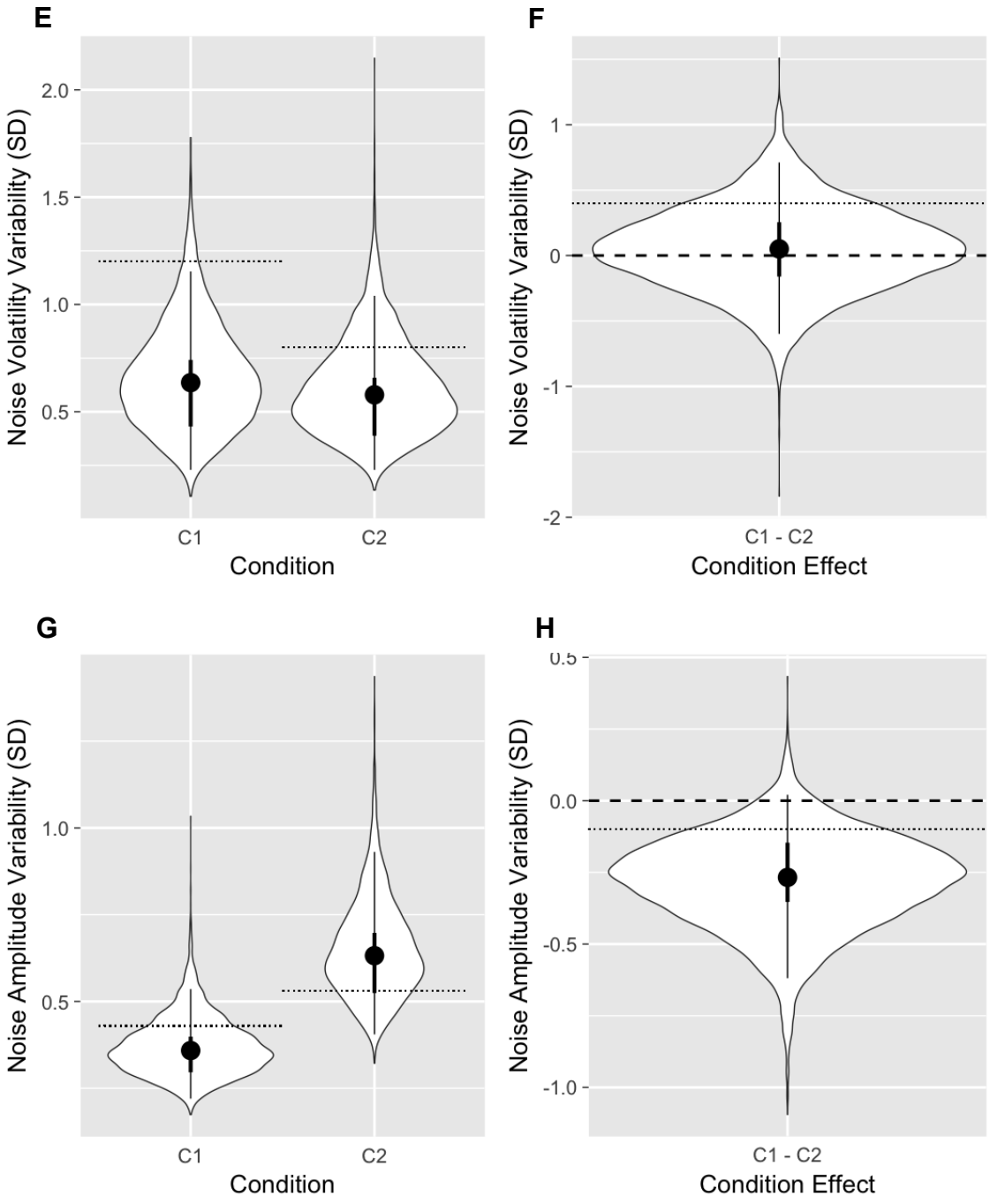

E
Noise Volatility Variability (SD)
2.0
1.5
1.0
0.5
C1
C2
Condition
F
Noise Volatility Variability (SD)
1
0
-1
-2
C1 - C2
Condition Effect
G
Noise Amplitude Variability (SD)
1.0
0.5
C1
C2
Condition
H
Noise Amplitude Variability (SD)
0.5
0.0
-0.5
-1.0
C1 - C2
Condition Effect

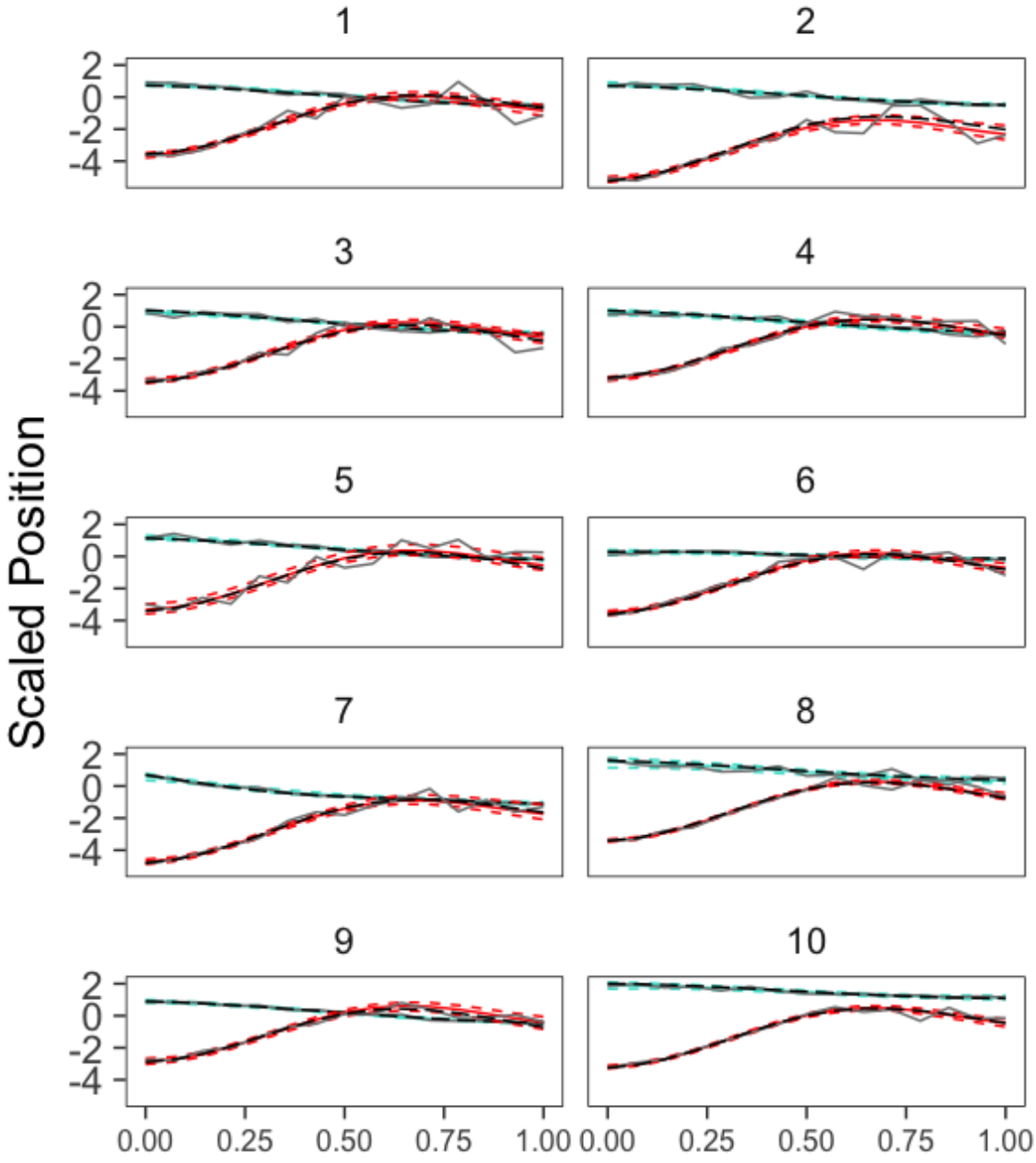
Figure 24 Violin plots of population noise volatility (A & B) and noise amplitude (C & D) estimates, and of population noise volatility (E & F) and noise amplitude (G & H) variability estimates. Posterior distributions are mirrored along the vertical axis for each estimate. For each row, the left graph displays the estimates as a function of condition whereas the right graph displays the effect of condition (condition 1 minus condition 2). The dashed black lines on the graphs on the right represents a null effect (zero). The horizontal dotted lines are the real hyperparameter values that were used to generate the simulated data. The large black dot of each estimate represents the median of the posterior distribution. The thick black lines represent 50% credible intervals and the thin black lines represent 95% credible intervals. Condition 1 and condition 2 are represented by "C1" and "C2", respectively.

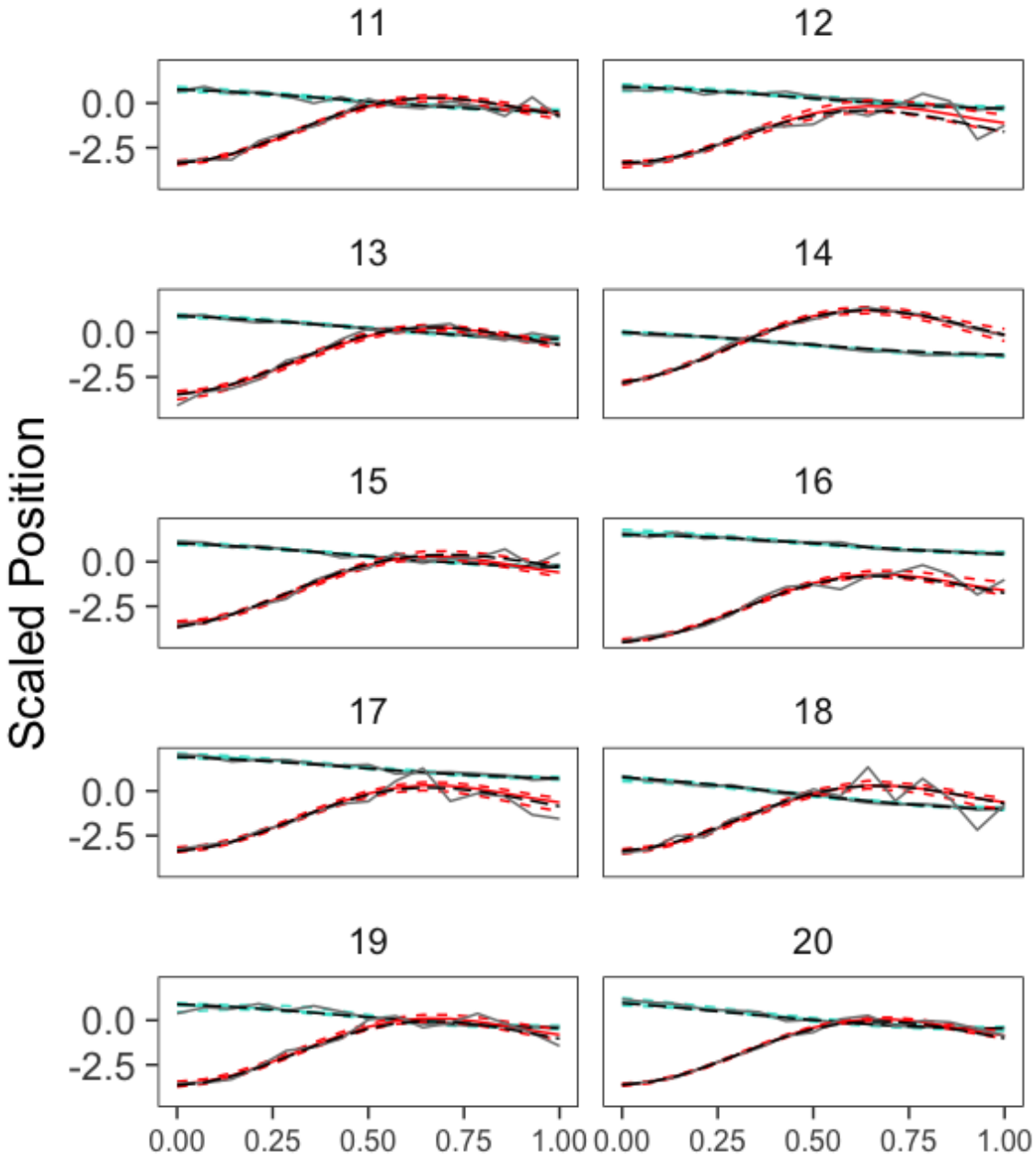
Noise:	Volatility	Amplitude	Volatility (SD)	Amplitude (SD)
C1	50%	95%	Excluded	95%
C2	50%	95%	95%	50%
Effect	50%	50%	95%	95%
Zero	95%	50%	50%	95%

Table 5 Summary of degree to which population noise hyperparameter estimates – noise volatility, noise amplitude, population noise volatility variability, and population noise amplitude variability - for condition 1 (C1), condition 2 (C2), and the effect (C1 minus C2) captured the true population hyperparameter values. If the true hyperparameter value was captured by the 50% credible interval, "50%" was entered into the corresponding box and the box was highlighted as green to convey that the population hyperparameter had been accurately estimated. If the true hyperparameter value was captured by only the 95% credible interval, "95%" was entered into the corresponding box and the box was highlighted as *light* green to convey that the population hyperparameter had been somewhat accurately estimated. Finally, if the true hyperparameter value was not captured by the 95% credible interval, "Excluded" was entered into the corresponding box and the box was highlighted as red to convey that the population hyperparameter had not been accurately estimated. Further, the bottom row of the table was added to indicate the degree to which the estimate of the effect of condition excluded zero (a null effect) in order to gauge the sort of conclusion that would have been drawn regarding the probability of the effect having had no knowledge of the true population effect (i.e. in an experimental setting). The coloring and entry scheme was reversed for this row of the table except in cases where there was truly no effect of condition (noise amplitude; see Figure 24, graphs C & D).

5.2.4.3 Participant Mean Functions

The participant mean function estimates are displayed in Figure 25. The participant-wise mean trajectory estimates captured the real participant mean functions. The estimates did not capture the empirical participant means particularly well (see participant 24, condition 2). However, the empirical participant means were noisy. Thus, the model managed to capture the true participant mean functions despite particularly noisy empirical data. Figure 26 displays the median values of the posterior estimates across participants for each condition to facilitate the detection of outliers. There were no notable outliers among the participant mean functions for either condition.


1
2
3
4
5
6
7
8
9
10
Scaled Position
2
0
-2
-4
0.00
0.25
0.50
0.75
1.00


11
12
13
14
15
16
17
18
19
20
Scaled Position
0.0
-2.5
0.00
0.25
0.50
0.75
1.00

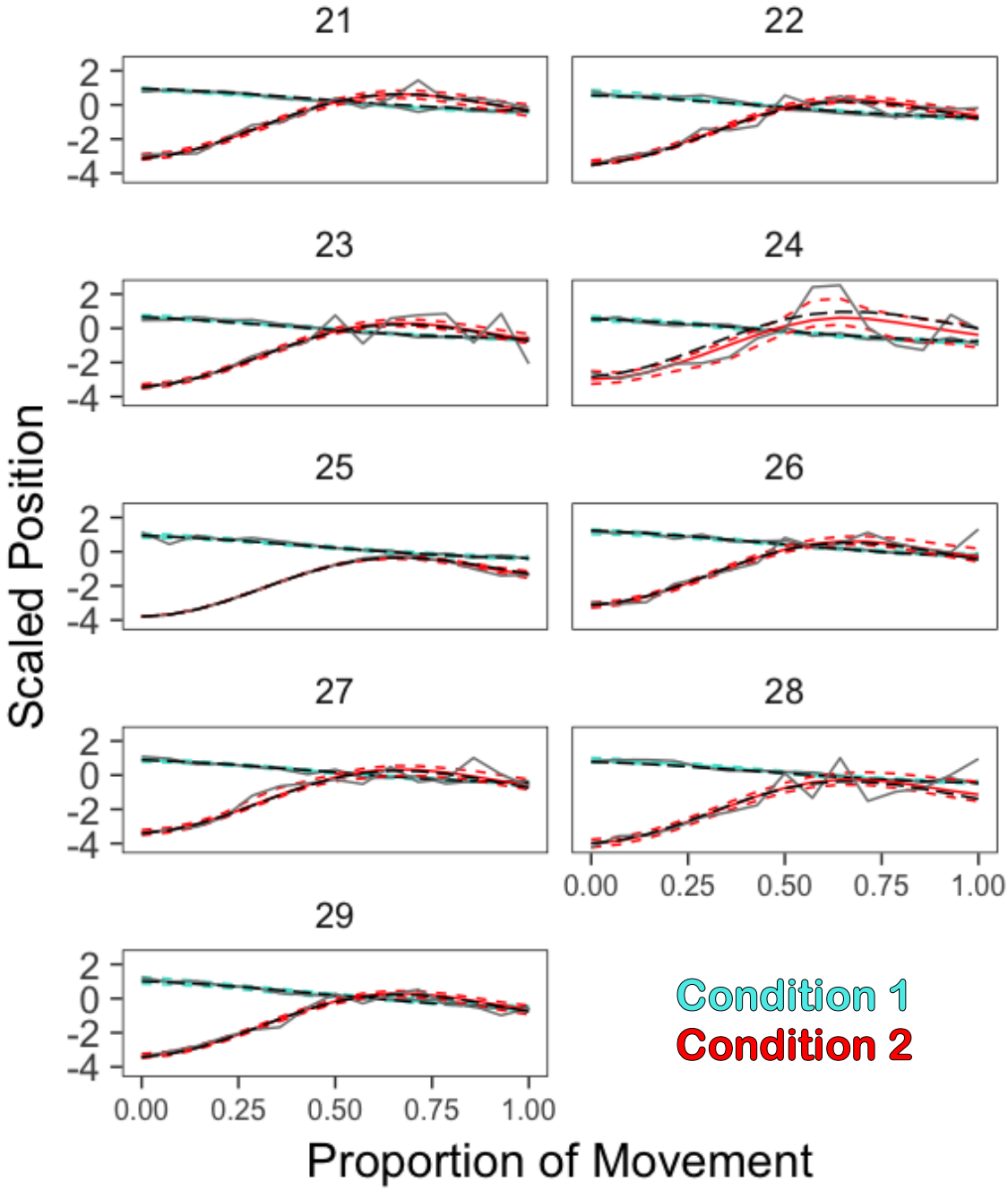

21
22
23
24
25
26
27
28
29
2
0
-2
-4
0.00
0.25
0.50
0.75
1.00
Scaled Position
Proportion of Movement
Condition 1
Condition 2

Figure 25 The mean function estimates for each participant, for conditions 1 (turquoise) and 2 (red), with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant mean functions. The dashed colored lines, surrounding the solid colored lines, make up the 95% credible intervals of the participant mean function estimates. The solid light gray lines represent the empirical means of the normalized data. The black dashed lines represent the real, latent, participant mean functions.

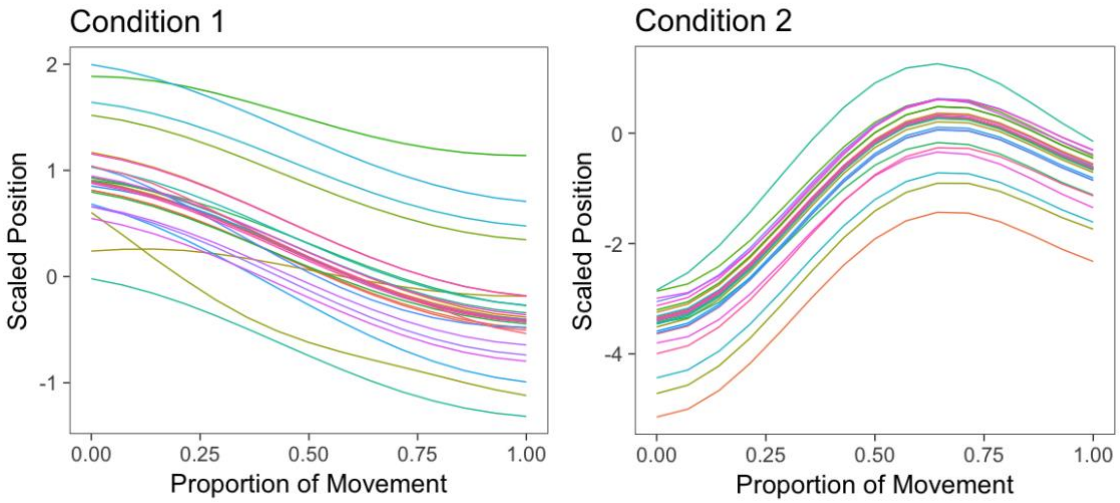
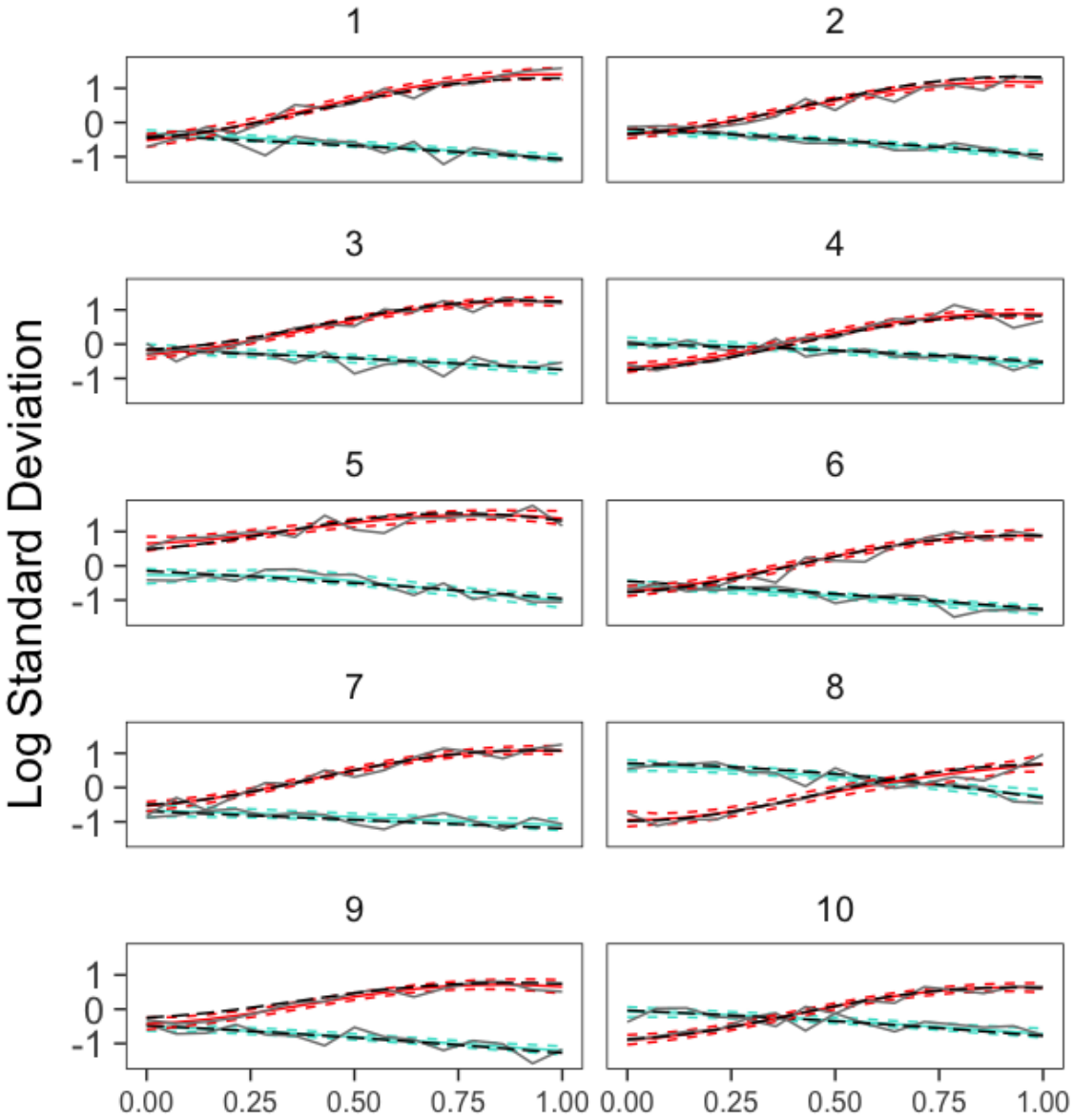


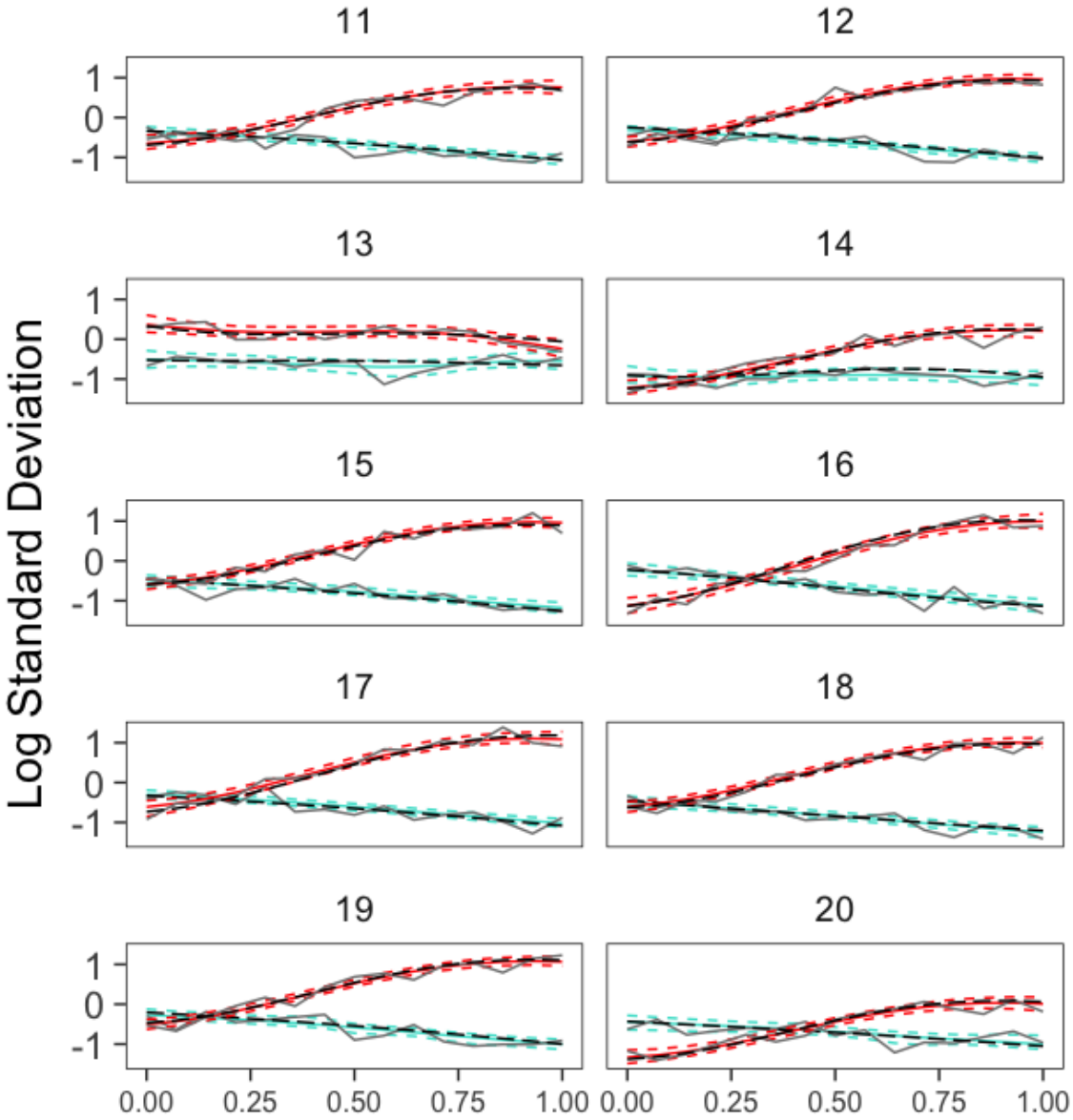
Figure 26 The mean function estimates for each participant, for conditions 1 and 2, with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant mean functions.

5.2.4.4 Hyperparameters for Noise Functions

The participant noise function estimates are displayed in Figure 27. The participant-wise noise function estimates captured the real participant mean functions. As with the participant mean functions, the estimates did not capture the empirical participant noise

functions particularly well because the empirical data were noisy. Again, the model managed to estimate the true participant functions despite noisy empirical data. Figure 28 displays the median values of the posterior estimates across participants for each condition to facilitate the detection of outliers. There was one notable outlier among the participant noise functions for condition 2. This participant had relatively little trial-by-trial variability early in their movement for condition 2.


Log Standard Deviation
1
2
3
4
5
6
7
8
9
10
1
0
-1
0.00
0.25
0.50
0.75
1.00


11
12
13
14
15
16
17
18
19
20
Log Standard Deviation
1
0
-1
0.00
0.25
0.50
0.75
1.00

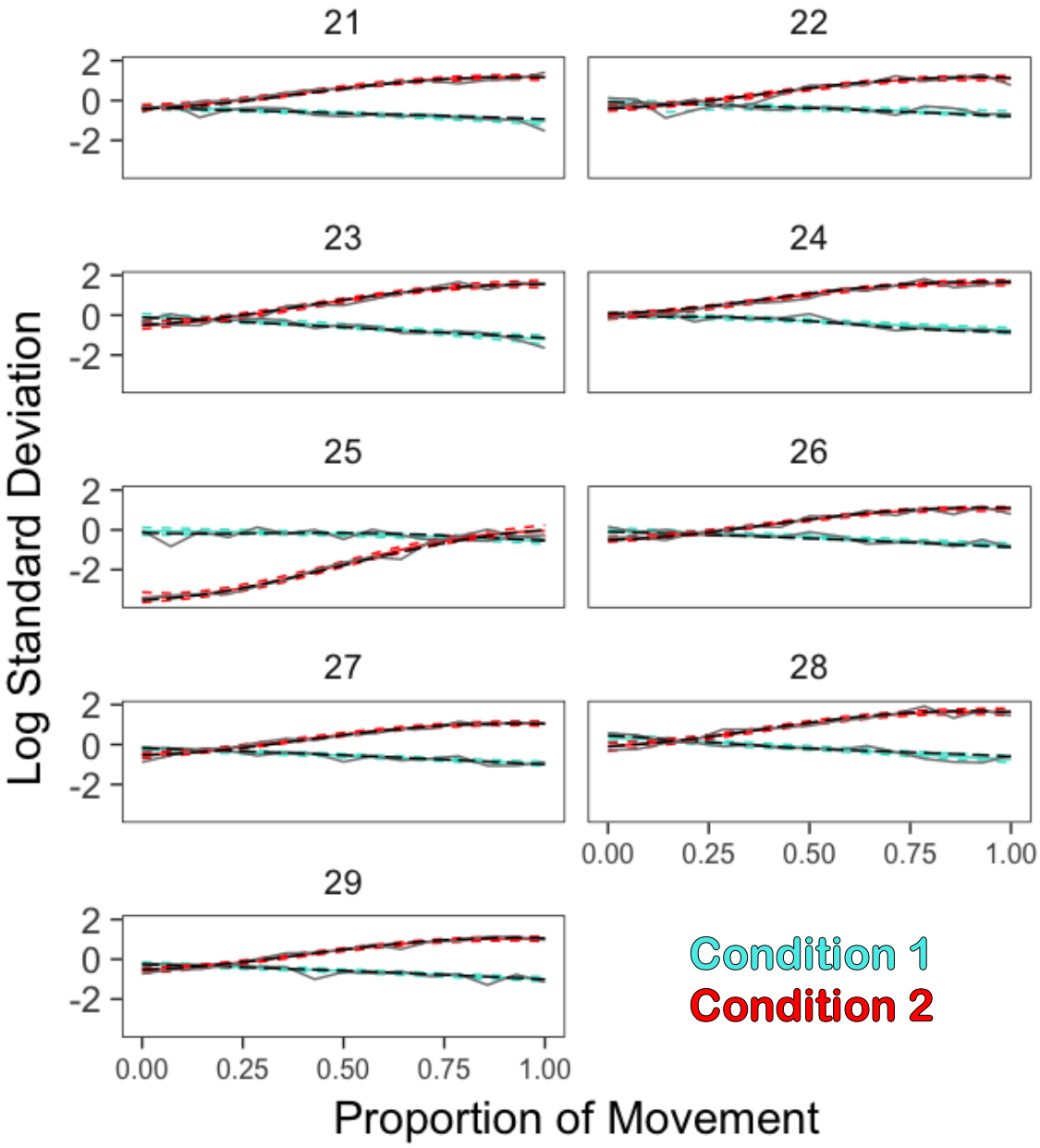

21
22
23
24
25
26
27
28
29
2
0
-2
0.00
0.25
0.50
0.75
1.00
Log Standard Deviation
Proportion of Movement
Condition 1
Condition 2

Figure 27 The noise function estimates for each participant, for conditions 1 (turquoise) and 2 (red) conditions, with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant noise functions. The dashed colored lines, surrounding the solid colored lines, make up the 95% credible intervals of the participant noise function estimates. The solid light gray lines represent the empirical variability (log SDs) of the normalized data. The black dashed lines represent the real, latent, participant noise functions.

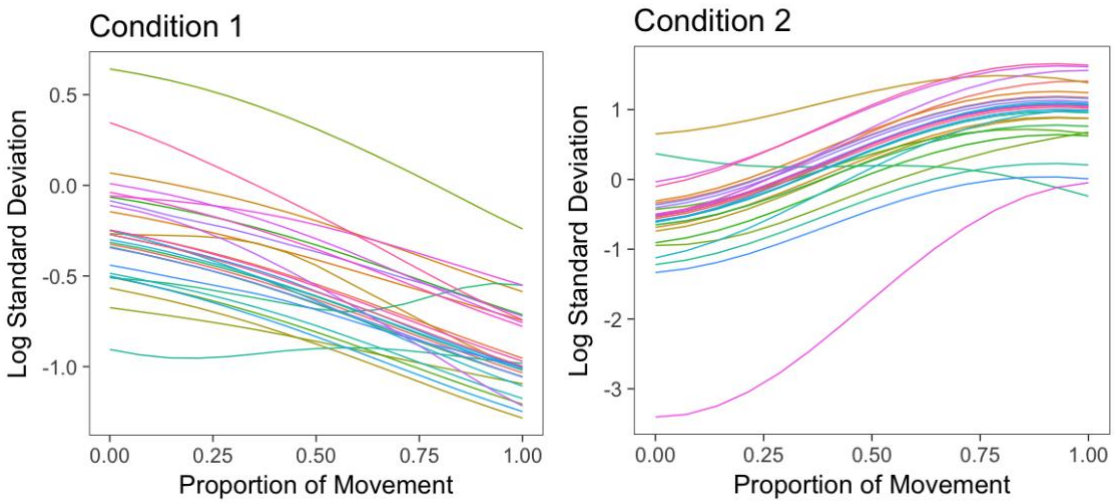


Figure 28 The noise function estimates for each participant, for conditions 1 and condition 2, with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates of the participant noise functions.

5.2.4.5 Population Mean Functions

The population mean function estimates are displayed in Figure 29. The population mean function estimates captured the true population mean functions. The estimates also tracked the empirical population mean functions fairly well, however these were noisy as

a result of the noisy participant-wise functions. There was large effect of condition on the population mean functions such that condition 1 had much greater position than condition 2 for the first half of the movement. This trend was reversed around three quarters way through the movement. These results were similar to those obtained from the fANOVA analysis. However, there was less uncertainty associated with the Bayesian estimates which allowed for the detection of an effect of condition around the 0.75 mark of the proportion of movement – an effect not detected by the fANOVA analysis. Furthermore, the fANOVA estimates slightly underestimated the population mean functions for condition 2, whereas the Bayesian estimates captured the true population mean function very well.

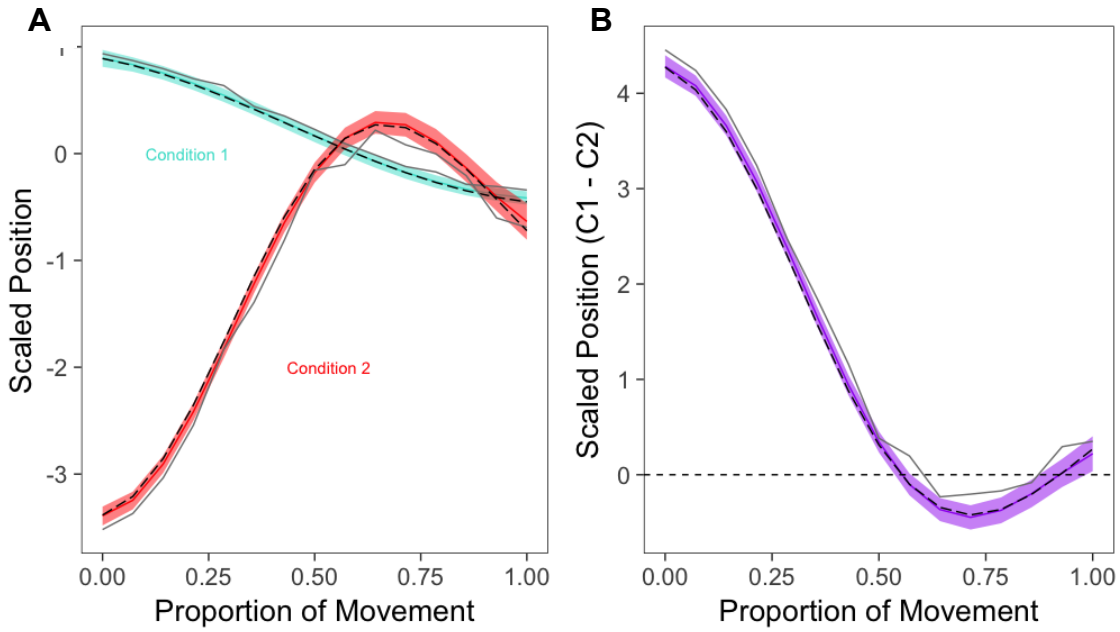


Figure 29 The population mean function estimates for condition 1 and condition 2 (A), and the effect of condition (condition 1 minus condition 2; B), with scaled position (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior. The error bands surrounding the solid colored lines make up the 95% credible intervals of the posterior estimates. The solid light gray lines represent the empirical means of the normalized data. The black dashed lines represent the real, latent, population mean functions. Condition 1 and condition 2 are represented by "C1" and "C2", respectively.

5.2.4.6 Population Noise Functions

The population noise function estimates are displayed in Figure 30. The population noise function estimate for condition 2 captured the true population mean function. However, the population noise function estimate underestimated the trial-wise variability for

condition 1. Unfortunately, it is unclear what the cause of this underestimation was since the other Bayesian model estimates were quite accurate. The estimate for condition 1 tracked the empirical population noise function fairly well. However, the estimate for condition 2 overestimated the variability compared to the empirical population noise function. The empirical population noise function for condition 2 was likely negatively biased by the outlying participant which had substantially smaller trial-wise variability values across the trajectory than the bulk of the participants (see Figure 24).

The estimate of the effect of condition slightly underestimated the true effect of condition. This was due to the underestimation of the true population noise function for condition 1. There was large effect of condition on the population noise functions such that condition 2 had much greater variability than condition 1 for the second half of the movement. The opposite trend was present to a lesser extent early in the movement. These results are largely consistent with those of the spatial variability profile analysis. However, the latter failed to capture the nuanced effect of condition on population noise very early in the movement (0.00 of proportion of time), whereas the Bayesian analysis detected this effect.

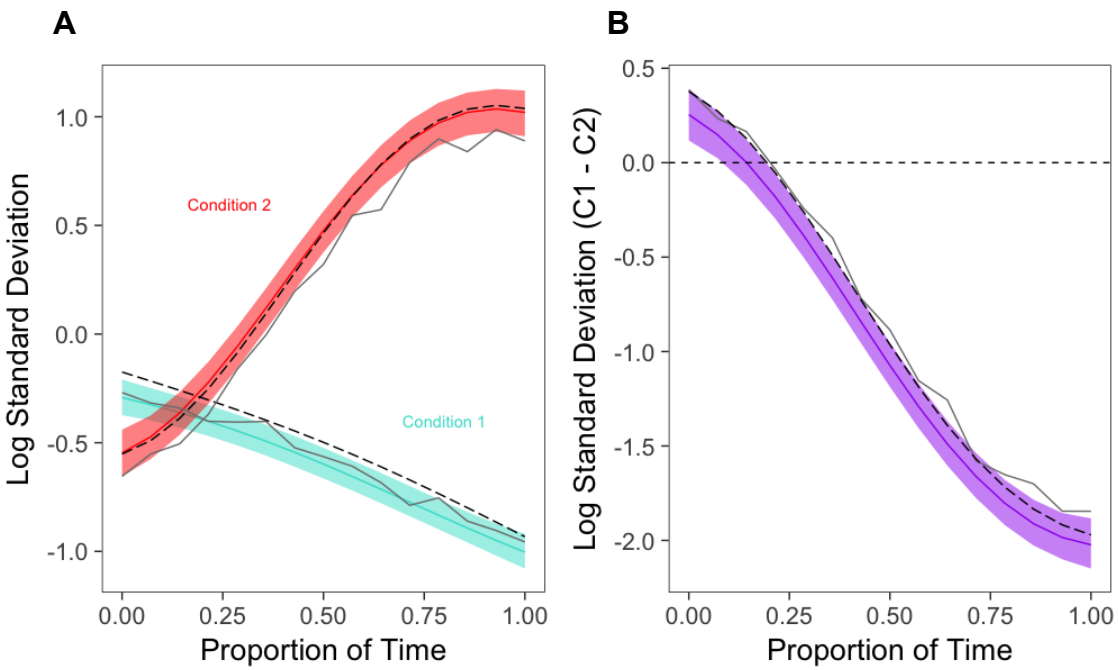


Figure 30 The population noise function estimates for condition 1 and condition 2 (A), and the effect of condition (condition 1 minus condition 2; B), with log standard deviation (y-axis) as a function of proportion of movement (x-axis). The solid colored lines represent the median of the posterior estimates. The error bands surrounding the solid colored lines make up the 95% credible intervals of the posterior estimates. The solid light gray lines represent the empirical variability (log SDs) of the normalized data. The black dashed lines represent the real, latent, population noise functions. Condition 1 and condition 2 are represented by "C1" and "C2", respectively.

5.3 CONCLUSION

R^2 analysis did not reveal any significant effects. The spatial variability profile analysis detected an interaction effect that was characterized by greater spatial variability in the condition 2 compared to condition 1, later in the movement. Further, the parameter estimates captured the true population noise function fairly well, although the series of estimates for condition 2 failed to discount the inherent noise in the simulated data and

therefore lacked some smoothness. The fANOVA estimates indicated that the position values of condition 1 were larger than those of condition 2, early in the movement. The estimates captured the true population mean functions well.

For the Bayesian hierarchical GPR analysis, broadly, the hyperparameter estimates for the population mean and noise functions captured the true hyperparameter values. The participant noise and mean function estimates captured the true participant functions despite substantial noisiness in the simulated dataset. The population mean function estimates captured the true functions well. Further, the estimates appropriately indicated an effect of condition later in the movement – this effect was not detected by the fANOVA analysis because of less precision in the estimates. The population noise function estimates captured the true functions well for condition 2, but not as well for condition 1. The cause of this inaccuracy is unclear. Regardless, the estimate of the effect of condition on population noise function captured the true function reasonably well and detected an early effect of vision. This early effect was not detected by the spatial variability profile analysis. Furthermore, it is unclear that it would be appropriate to make specific pairwise comparisons in this manner when interpreting the spatial variability profile results, without accounting for multiple comparisons.

CHAPTER 6 DISCUSSION

6.1 OBJECTIVES

The primary objective of the present thesis was to propose and assess a new method of analyzing trajectory data in the context of goal-directed reaching. The proposed method, Bayesian hierarchical GPR, was compared to statistical methods that have been traditionally used in the field, and more novel methods that have been proposed in the past few years. The authors believe this to be the first comprehensive application of Bayesian hierarchical GPR to movement trajectories in the context of online motor control and goal-directed reaching tasks.

The secondary objective was to partially replicate the results from the basic goal-directed reaching task implemented by Elliott & Hansen (2010), which involved removing vision during movement for some trials. Elliot & Hansen (2010) compared multiple traditional analysis methods with respect to their ability to detect the effect of vision. In the current study, the relative effect sizes and the magnitude of the effects from the traditional analyses were compared to those of Elliott & Hansen (2010).

6.2 METHODS

Several statistical analyses were applied to a dataset derived from a basic goal-directed reaching task. Participants reached towards a target as quickly and accurately as possible. In a blocked design, in half the trials, participant vision was removed upon movement onset and only regained upon movement offset. A gamut of statistical analyses was performed on the corresponding trajectory data. Among these analysis methods, was the proposed Bayesian hierarchical GPR method which primarily estimated the population mean function and noise functions for the trajectory amplitudes. For the cases in which it was meaningful to do so, the traditional statistical methods were compared to the proposed Bayesian method with respect to their ability to detect effects generated by the experimental manipulation (vision vs. no-vision) and the nuance with which the estimates captured these effects.

In addition, several of the statistical methods that were applied to the experimental dataset were also applied to simulated dataset. Arbitrary functions were simulated from a hierarchical generative model identical to the analytical GPR model. Due to their arbitrary nature, the simulated functions were not necessarily representative of actual movement trajectories that one would expect to collect from a goal-directed reaching task. However, since the data were simulated, the population mean and noise functions (and the other parameters of the generative model) were known and could therefore be compared to the estimates generated by the list of applicable statistical methods. This allowed the authors to directly assess the accuracy of the estimates across statistical methods.

6.3 SUMMARY OF RESULTS

Largely, the expected experimental effects were generated. The experimental manipulation of participant vision affected the nature of the movement trajectories. As expected, with access to vision, participants initiated their movements more rapidly and achieved peak velocity proportionally earlier in their movements. Further, participants were more accurate when vision was available during their movements. Finally, early in the movement, trial-by-trial movement variability was greater when vision was available. Similar to what was discovered by Elliot & Hansen (2010), the analysis of the spatial variability profile method detected the largest vision condition effects and were therefore purported to be among the more sensitive traditional analysis methods. Both the fANOVA (Gallivan & Chapman, 2014) and the Bayesian hierarchical GPR analyses detected notable differences in the movement trajectory amplitudes. The latter also detected notable differences between the population noise functions of the two vision conditions, mirroring the spatial variability profile results.

Although several traditional methods successfully detected vision condition effects, many of them lacked severable desirable properties, both in isolation and in combination. For instance, the outcome DVs – RT, MT, Target Error, VE, CE – were analyzed separately. However, it is reasonable to assume that RT and MT may covary across participants, in the same way that the error-related DVs may well covary. The statistical tests applied to

the outcome variables in the current thesis, which were made to be similar to those that are commonly applied in the field, did not account for this covariance structure. This omission may have negatively affected proper estimation of effect parameters and failed to give the researchers the opportunity to examine meaningful correlation among the outcome variables. Another serious issue with the application of separate rANOVAs across the outcome variables is the inflation of Type I error rates. If one were studying a new phenomenon and had no *a priori* expectations for each the outcome variables, then would one conclude that just one significant result among all the possible significant results indicates the success of the experimental manipulation? Further, if one applied multiple comparison corrections across tests, would one fail to detect meaningful effects (i.e. inflation of type II error)? The issue, ultimately, is that the series of rANOVAs applied in the current thesis do not comprehensively model the group of outcome variables.

The proportional TAPV analysis detected a small effect of vision. However, it did not provide substantial information and nuance about the nature of the effect under examination. Although it is useful to know that the kinematic structure of two sets of trajectories differ between experimental conditions, it is more useful to understand the way in which they differ. Indeed, the fANOVA and GPR methods provided normalized estimates of the population mean functions. The information regarding TAPV was embedded in these results. Further, the function estimates provided more detailed accounts of how the trajectories were affected by the experimental manipulation.

The discontinuity analysis measured a relatively small effect size for the effect of vision. Indeed, although this method may have been useful in the current experimental paradigm in which discontinuity counts differed between conditions, it has the inherent shortcoming of not providing any detail about differences between the shapes of trajectories across conditions and cannot detect adjustments that may be made in a more continuous or smooth manner. The R^2 analysis suffered from the same issue in principle: it only provided a crude measure of auto-correlation at certain points in the movement. In the present thesis, the R^2 analysis also failed to detect the expected interaction effect.

The spatial variability and epsilloid methods both illustrated the manner in which trial-wise variability differed between conditions in reasonable details. The epsilloid method offered the comparative advantage of having greater temporal resolution and accounting for all three dimensions of the movement. The issue with both these methods, however, is that they only examine movement variability across trials and do not account for differences in the actual trajectories. Furthermore, for both methods, it is unclear how inferences are to be made regarding trial-wise variability differences at specific normalized time points along the spatial variability profile. The multiple comparison issues that have been addressed throughout the current thesis come up again when trying to make inferences beyond the trend-based interpretation of the rANOVAs for these spatial variability methods.

The fANOVA analysis provided a nuanced representation of the manner in which the trajectory amplitudes differed between conditions. Unfortunately, the fANOVA method comes with its own interpretive issues. For instance, in the current thesis, the fANOVA analysis effectively consisted of a series of t-tests, one for each time point. Gallivan & Chapman (2014) argue that the fANOVA ought to be interpreted as a single data object, thereby making inference statements regarding trajectory differences at any particular point in time verboten. It then becomes unclear what inferential value is added by the fANOVA. Does the fANOVA reduce itself to a descriptive statistical tool as opposed to an inferential one? How precisely is one meant to comment on a series of t-tests results in a holistic manner? Ultimately, the precise inference statements one ought to make about an fANOVA object are unclear. Furthermore, the fANOVA analysis suffers from a shortcoming that is shared among the other traditional analysis methods: it does not model the data hierarchically and therefore fails to account for the uncertainty in the data at the participant level. The following section will discuss the ways in which the Bayesian hierarchical GPR method addresses these and other issues discussed in this subsection.

6.4 GPR vs. Traditional Analyses

The GPR analysis outperformed the other analysis methods in several respects. For one, the GPR analysis provided a more comprehensive account of the results. This was largely a result of the nature of the model itself. The Bayesian hierarchical GPR model provided estimates for participant mean and noise functions, and for population mean and noise functions and their hyperparameters. Therefore, whereas the fANOVA method provided

estimates for population movement amplitude and the spatial variability profile analysis provided estimates for the population noise function, the GPR model did both under a single cohesive statistical framework.

Another advantage to the GPR method is that it tended to produce estimates with less uncertainty (i.e. tighter error bars). Indeed, the error bars of the GPR population mean function estimates for the simulated dataset were notably tighter than those for of the fANOVA estimates. For the experimental dataset, the population noise function estimates of the GPR model captured difference between conditions later in the movement that were not detected by the epsilloid or spatial variability profile methods.

More broadly, it is not even clear whether the error bars that were selected for the epsilloid method were appropriate for evaluating differences between the conditions along the movement trajectory. The epsilloid method primarily provided an interaction term which does not provide inferential nuance regarding the manner in which the interaction effect instantiated itself in the trial-wise variability data. The interpretation of the GPR method is much more explicit: if the estimate of the effect of condition on population noise function excludes zero at given time points, then the conditions credibly differ at those time points in terms of variability. Furthermore, the spatial variability profile analysis, although it provides estimates at key points throughout the movement (kinematic markers), lacks a certain degree of temporal resolution and therefore nuance regarding the nature of the underlying population noise functions. In contrast, the GPR

method allows the researcher to examine differences in trial-wise variability between conditions at multiple (normalized) time points throughout the movement.

The GPR analysis was not directly comparable to all the traditional analyses. For instance, looking at DVs such as RT, MT, and Target Error still offers some unique insights about how participants performed the task. The GPR estimates convey inferential information about the nature of the trajectories themselves – the shape of the trajectories and how the trajectories vary across trials. The GPR model does not, however, confer any insights regarding the absolute time it took participants to complete their movements or the accuracy with which the participants reached the target. In theory, DVs such as MT and Target Error could be incorporated into the Bayesian hierarchical GPR model. However, in practice, such a model would not be computationally tractable (for discussion, see subsection 6.5). One could design a separate Bayesian hierarchical model to deal specifically for DVs such as MT and Target Error, and to evaluate the extent to which these covary. However, such an endeavor is beyond the scope of the current thesis, which is centered on the application of GPRs to trajectory data.

6.5 LIMITATIONS

Although it is believed that the Bayesian hierarchical GPR model proposed herein provided substantial benefits to the analysis of limb trajectory data in the context of visual control, the model was far from ideal. For instance, the model only assessed one of

the three dimensions of space (i.e. movement amplitude). Ideally, the model would account for all three dimensions of space and relate them through time. With the proposed model, the model would have to be run separately in each of three dimensions.

Another noteworthy limitation of the proposed GPR model is that it fails to fully account for trial-by-trial deviations in a participant's trajectory. Indeed, the model lacks a trial-level hierarchy in its structure. One consequence of this shortcoming is that when the model is used to generate samples (for a simulation), trial-wise trajectories are not represented as smooth continuous functions that are slightly altered in shape and position. Instead, trial-wise trajectories are represented as noisy (caterpillar-looking) functions. The failure to model each trial appropriately may result in loss precision in the participant-wise trajectory estimates.

There are also general concerns about the normalization procedure implemented herein. Gallivan & Chapman (2014) argued that normalization with respect to a dimension (time or space) that varies across conditions can introduce misleading artifacts. In this particular case, these concerns are assuaged: as with Elliott & Hansen (2010), MT differences between the two vision conditions were not observed, suggesting that the trajectories did not differ substantially with respect to the dimension of time. Therefore, it is assessed that the normalization procedure likely did not introduce misleading artifacts. The current analysis involved a time normalization because the shape of the trajectory as a function of time was of interest. Future analyses, however, could be normalized to

space.

Some of the standard model validation procedures for Bayesian models were not implemented. Indeed, there are multiple ways to validate Bayesian models beyond basic checks of Rhat and effective sample size (for discussion, see Gelman et al., 2014; and Stan Development Team, 2017). For instance, one can use posterior predictive checks (PPCs), which involve simulating data from the posterior distribution and evaluating the extent to which those simulated values match the raw data. If the model is misspecified, then the simulated values will not adequately resemble the raw data. Although, PPCs were not conducted for the proposed Bayesian model, the raw data were plotted along with the Bayesian estimates throughout the thesis to give the reader a visualization of how the model estimates matched the raw data. The results were deemed to be satisfactory because model estimates matched the raw data in most cases.

Sensitivity analyses, which involve running the model with various plausible probability models (in the prior or the likelihood) to evaluate the extent to which the resulting posterior distribution is sensitive to the choice of probability models, constitute another validation method. This method of validation was run informally during the model building stage of the current thesis. Multiple prior distributions were tested, and they did not appear to have a substantial impact on the resulting posterior estimates. Indeed, this finding was not surprising given that the priors were selected to be uninformative, thereby giving more weight to the actual data in determining the posterior distributions. Importantly, a full-scale simulation was run demonstrating that the Bayesian estimates reliably captured the true underlying population and participant parameters of the

generative model, providing a strong validation of the model.

Unfortunately, no quantitative method for evaluating the results of the simulation were used. Instead, the extent to which the estimates from the various statistical methods captured the underlying population and participant parameters was assessed via visual inspection. Therefore, the assessment of the relative accuracy of the various statistical methods was inherently subjective. The statistical outputs for the experimental dataset also lacked an objective evaluative framework. Indeed, although effect sizes were used to compare model sensitivity among basic frequentist methods, they cannot be compared with more complex frequentist methods or with the Bayesian method introduced in the current thesis.

Ideally, a single parameter would be used to compare the quality ('goodness of fit' and predictive ability) of several models. In the frequentist realm, when the generative model is unknown, AIC scores are often used to compare the quality of different models (for discussion, see Burnham, Anderson, and Huyvaert, 2011). In the Bayesian realm, Bayes Factors are often used to compare the quality of different models (for discussion, see Kruschke, 2014). Unfortunately, it is unclear what metric could be used to evaluate the quality of statistical models across frequentist and Bayesian lines. Moreover, it is unclear that it would make sense to directly compare results from these two statistical paradigms. Finally, it is unknown which metric would be appropriate for evaluating accuracy of estimates from different models when the true parameter values are known (i.e. in the case of a simulation study).

Another limitation was alluded to in subsection 4.1: the trajectories must be downsampled in order to make the analysis computationally tractable. In theory, the negative consequence of this limitation should only substantively manifest itself if the trajectories are reasonably complicated. The experimental design presented herein generated simple movement trajectories. Therefore, downsampling should not have had a substantial negative impact on the analysis. Indeed, comparisons between the GPR results and those of the fANOVA and spatial variability analyses did not indicate that downsampling had an adverse effect on the results. Of course, issues stemming from downsampling would potentially arise if the GPR analysis were used to model trajectory estimates with multiple discrete, rapid, and subtle error-corrections.

Ultimately, the downsampling issue discussed above is a consequence of a broader issue of computational capacity. As stated in subsection 4.2, the Bayesian hierarchical analysis of the experimental data took approximately 120 hours to run. Ideally, even reasonably sophisticated analyses should be implementable overnight. The requirement for several days of processing has serious practical implications. For example, it limits the researcher's ability to make minor changes to the model structure and get results in a reasonable amount of time, thereby hampering the iterative process of model fine-tuning. The time constraint would also surely discourage many researchers from even attempting to implement the model in the first place. The computational demands of the Bayesian hierarchical GPR model necessitates not only time but also access to reasonably powerful computers that can run the model on several cores and can be left running for several days.

The Bayesian hierarchical GPR model presents other usability issues. Namely, although the model estimates are straightforward and intuitive to interpret, the model structure itself is reasonably complex. A researcher would have to be familiar with prior distributions, hierarchical model structures, and GPs more broadly to intelligently construct a Bayesian hierarchical GPR model. These statistical concepts are not part of the typical researcher's knowledge base in the field of goal-directed reaching. Thus, most researchers would have to spend time getting familiar with these statistical concepts to implement the statistical model proposed in the current thesis. Luckily, Bayesian methods are gaining traction (Kruschke & Liddell, 2015), and therefore will become more accessible to those researchers that are interested in implementing Bayesian statistics in their work.

6.6 CONCLUSION

The Bayesian hierarchical GPR model was successfully implemented and conferred some substantial benefits with respect to its ability to generate a comprehensive set of accurate, precise, and relevant posterior estimates. Furthermore, the experimental manipulation largely produced the expected results, thereby partially replicating the results of Elliot & Hansen (2010). Despite these positive outcomes, the typical researcher may be hesitant to consider using this method because of the limitations listed in subsection 6.5. Ultimately, the complexity of the Bayesian hierarchical GPR model and its computational limitations may not outweigh its stated benefits. On the other hand, researchers with unique statistical backgrounds should endeavor to use the method because many of the stated

limitations would naturally become obsolete. Therefore, Bayesian hierarchical GPR should be considered a specialized tool until steps are taken to facilitate the implementation of the method.

BIBLIOGRAPHY

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. Eprint arXiv:1406.5823, 67(1), 51. http://doi.org/10.18637/jss.v067.i01

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behavioral Ecology and Sociobiology, 65(1), 23–35. http://doi.org/10.1007/s00265-010-1029-6

Cox, G. E., Kachergis, G., & Shiffrin, R. M. (2012). Gaussian Process Regression for Trajectory Analysis.

Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *The American Psychologist*, *60*(2), 170–180. http://doi.org/10.1037/0003-066X.60.2.170

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perpectives on Psychological Science*.

Elliott, D., Helsen, W. F., & Chua, R. (2001). A Century Later : Woodworth's (1899) Two-Component Model of Goal-Directed Aiming. *Psychological Bulletin*, *127*(3), 342–357. http://doi.org/10.1037//0033-2909.127.3.342

Elliott, D., & Hansen, S. (2010). Visual regulation of manual aiming: a comparison of methods. *Behavior Research Methods*, *42*(4), 1087–1095. http://doi.org/10.3758/BRM.42.4.1087

Elliott, D., Lyons, J., Hayes, S. J., Burkitt, J. J., Roberts, J. W., Grierson, L. E. M., …

Bennett, S. J. (2017). Neuroscience and Biobehavioral Reviews The multiple process model of goal-directed reaching revisited. *Neuroscience and Biobehavioral Reviews*, *72*, 95–110. http://doi.org/10.1016/j.neubiorev.2016.11.016

Flaxman, S., Gelman, A., Neill, D., & Wilson, A. G. (in preparation). Fast hierarchical Gaussian processes, 1–18.

Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in Neuroscience*, *8*(8 JUL), 1–19. http://doi.org/10.3389/fnins.2014.00215

Gelman, A., Carlin J. B., Stern, H.S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis.* Boca Ranton, Florida: CRC Press.

Grosbois, J. De, & Tremblay, L. (2016). Quantifying online visuomotor feedback utilization in the frequency domain. *Behavior Research Methods*, 1653–1666. http://doi.org/10.3758/s13428-015-0682-0

Hansen, S., Glazebrook, C. M., Anson, J. G., Weeks, D. J., & Law, H. (2006). The Influence of Advance Information About Target Location and Visual Feedback on Movement Planning and Execution. Canadian Journal of Experimental Psychology, 60(3), 200–208. http://doi.org/10.1037/cjep2006019

Hansen, S., Elliott, D., & Khan, M. (2007). Comparing derived and acquired acceleration profiles : 3-D optical electronic data analyses. *Behavior Research Methods*, *39*(4), 748–754.

Hansen, S., Elliott, D., & Khan, M. A. (2008). Quantifying the Variability of

Three-Dimensional Aiming Movements Using Ellipsoids. *Motor Control*, 241–251.

Heath, M., Westwood, D., & Binsted, G. (2004). The Control of Memory-Guided Reaching Movements in Peripersonal Space.

Heath, M. (2005). Role of limb and target vision in the online control of memory-guided reaches. *Motor Control*, *9*, 281–311.

Keele, S. W., & Posner, M. I. (1968). Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, *77*(1).

Khan, M. a, Chua, R., Elliott, D., Coull, J., & Lyons, J. (2002). Optimal Control Strategies Under Different Feedback Schedules : Kinematic Evidence. *Journal of Motor Behavior*, *34*(1), 45–57.

Khan, M. A., Franks, I. M., Elliott, D., Lawrence, G. P., Chua, R., Bernier, P. M., … Weeks, D. J. (2006). Inferring online and offline processing of visual feedback in target-directed movements from kinematic data. *Neuroscience and Biobehavioral Reviews*, *30*(8), 1106–1121. http://doi.org/10.1016/j.neubiorev.2006.05.002

Kruschke, J. K. (2012). Bayesian Estimation Supersedes the t Test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. http://doi.org/10.1037/a0029146

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press.

Kruschke, J. K., & Liddell, T. M. (2016). The Bayesian New Statistics : Hypothesis

testing , estimation , meta-analysis , and planning from a Bayesian perspective. *Psychonomic Bulletin & Review*, 1–53. http://doi.org/http://dx.doi.org/10.2139/ssrn.2606016

McElreath, R. (2015). Statistical rethinking: *A Bayesian course with examples in R and Stan* (Vol. 122). CRC Press.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. Psychological Methods, 8(4), 434–447. http://doi.org/10.1037/1082-989X.8.4.434

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rasmussen, C.E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Massachusetts: MIT Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*, 1359–1366. http://doi.org/10.1177/0956797611417632

Stan Development Team (2017). *Stan Modeling Language: User's Guide and Reference Manual.* Version 2.15.0.

Whitney, D., Westwood, D. a, & Goodale, M. a. (2003). The influence of visual motion on fast reaching movements to a stationary object. *Nature*, *423*(6942), 869–873. http://doi.org/10.1038/nature01693

APPENDIX A Consent Form

Project title: Measurement of Reach Trajectories with and without Visual Feedback

Lead researcher:
Mr. Ghislain d'Entremont
Master of Science Candidate
School of Health and Human Performance
Phone: (902) 802-5671
Email: gh679509@dal.ca
Supervisor:
Dr. Heather Neyedli
School of Health and Human Performance
Email: hneyedli@dal.ca
Phone: 902 494 6786

Funding provided by: National Science and Engineering Research Counsel of Canada (NSERC)

Introduction

We invite you to take part in a research study being conducted by Ghislain d'Entremont at Dalhousie University

The Health Sciences Research Ethics Board of Dalhousie University has reviewed the project and found it to conform to current ethical guidelines. It is the responsibility of us, the researchers, to adhere to these guidelines. These guidelines require:

1) That you be informed of the purpose of the research project and any attendant inconvenience, risk or benefits.
2) That the character of the task required be explained to you
   That you be made aware that participation is entirely your choice and that you may decline to continue at any point during the course of the research project, without loss of expected compensation. Further, if you are a student at Dalhousie University, there will not be any academic impact based on your decision of whether or not to participate and/or a decision to withdraw from the study at any point.

Please ask as many questions as you like. If you have questions later, please contact the lead researcher.

Purpose and Outline of the Research Study

Our study serves to characterize the movement pattern of the hand when reaching to targets with and without visual feedback. You will be asked to attend a single session lasting approximately 30 minutes.

Who Can Take Part in the Research Study

You can take part in this study if you are between 18-34 years of age, are right handed and have normal or corrected-to-normal vision. You have normal vision, as far as we are concerned, if you can read this sentence while holding this paper at arm's length.

What You Will Be Asked to Do

You will use a touch-screen computer screen, similar to that of a tablet. Infra-red emitting diodes (IREDs), which are small sensors about the size of a lentil, will be placed on your index finger, and the Optotrak Certus system will be used to collect positional data of your movement. (Note that the camera system only measures infra-red light, thus it only measures the position of the IRED and does not record images of you.) The position of the your index finger in 3D space will

be recorded in each trial.
There will be a 'home' location where you will be asked to keep their (right index) finger. Upon the onset of the trial the target will be appear after a variable period of time, signaling that you may begin your movement to it. You will be asked to reach and touch the target, which will appear as a small shape on the touch screen, as quickly and accurately as possible.
You will be asked to wear occlusion goggles throughout the duration of the experiment. Occlusion goggles are simply goggles that can be programmed to go from a see-through to an opaque state, thus eliminating your vision. On some trials, the occlusion goggles will shift to their opaque state upon movement initiation. Therefore, you will sometimes be making reaching movements to targets without being able to see your limb or the target.
The entire experiment will be broken up into blocks. You will complete 20 to 40 trials in each block.

Possible Benefits, Risks and Discomforts

There are no direct benefits to participating in this study. Participating in the study might not benefit you, but we may gain insight into the cognitive processes that are responsible for the online control of limb trajectories.
Risks: There are minimal risks for participating in this study and they are similar to working with a tablet or iPad. You may experience some mental fatigue while completing the response selection task. Breaks will be provided in order to combat mental fatigue.

Compensation / Reimbursement

To thank you for your time compensation will be awarded in the form of 0.5 credits per half hour (if recruited through SONA).

How your information will be protected:

All consent forms from the study will be kept in the Cognitive Motor Performance Lab (Dr. Heather Neyedli). The lab will remain locked when no other lab members are within it. Therefore, only lab members will have access to the consent forms.
We will describe and share our findings in papers and/or presentations. We will be very careful that no one will be identified. This means that ***you will not be identified in any way in our reports***. The people who work with us have an obligation to keep all research information private. Also, we will use a participant number (not your name) in our written and computer records so that the information we have about you contains no names. The data we recorded from your movements may be made available to the public, in which case your participant number will be randomized so that your order of participation (e.g. were you the first participant?) cannot be inferred. In particular, your data will be made available via a software sharing service such as GitHub and/or via an R package. The basic demographic information we collect at the beginning of the experiment (age and sex) will **not** be made available.

If You Decide to Stop Participating

You are free to leave the study at any time. If you decide to stop participating at any point in the study, you can also decide whether you want any of the information that you have contributed up to that point to be removed or you will allow us to use that information. You can also ask for your data to be removed immediately after the completion of the study. However, once you leave research laboratory following the completion of the study, you will no longer be able to remove your data. The reason for this limitation is that we want the ability to publish participant data online for the sake shared analysis and storage throughout the course of the data collection process.
It is important to make the data publically available given that this is an analysis based project. This will allow other researchers to replicate our analyses and compare the results obtained from other analysis techniques beyond the techniques used in the present study.

How to Obtain Results

If you would like a copy of the study results please contact Ghislain d'Entremont (902-802-5671 or ghislaindentremont@gmail.com)

Questions
If you have questions or concerns about your participation in this research study please contact Ghislain d'Entremont at any time with questions, comments, or concerns about the research study.
If you have any ethical concerns about your participation in this research, you may also contact Research Ethics, Dalhousie University at (902) 494-1462, or email: ethics@dal.ca (and reference REB file # 2017-4239).

DALHOUSIE UNIVERSITY
Inspiring Minds

Signature Page

Project title: Measurement of Reach Trajectories with and without Visual Feedback
Lead researcher:
Mr. Ghislain d'Entremont
Master of Science Candidate
School of Health and Human Performance
Phone: (902) 802-5671
Email: ghislaindentremont@gmail.com

I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I understand that I have been asked to take part in one session in the Cognitive and Motor Performance Lab at Dalhousie University. I agree to take part in this study. My participation is voluntary and I understand that I am free to withdraw from the study at any time, and my data can be withdrawn up until the time I leave the research laboratory. I understand that my data may be made available to the public, but that a participant number will be used instead of my name to maintain the anonymity of my data.

_____ _____ _____

Name Signature Date

☐ **For SONA participants, please check this box if you wish to participate as an observer. As an observer, your data will not be used in the study as discarded upon completion of your participation.**

APPENDIX B Verbal Consent Form Clarification

You have just read through the consent form. Now, I would like to clarify a few things about the study before you give your written consent. First, I will briefly explain the design. You will make speeded reaching movements towards visual targets as they appear on a touch screen. Sometimes, you will lose your vision during your movement because the goggles you will be wearing will become opaque.

Now, I will clarify what we will do with your data after data collection. Information about your movements trajectories and responses will be made publicly available. However, your identify will in **not** be associated with the data. Only a randomized participant ID will be associated with your data.

APPENDIX C Participant Instructions

You will complete two experimental blocks. Each block will consist of 20 trials. Before each block there will be a practice block of 10 trials. You will be wearing goggles that will sometimes block your vision during your movement.

In each trial, you will be required to make a reaching movement towards a visual target. The trial will begin once you press the cross at the bottom of the touch screen, at which point you will hear a brief 'beep' sound. Shortly after, a target (a small white square) will appear on the screen. You will be required to touch the target with your right index finger as quickly and as accurately as possible.

In one of the two experimental blocks, the goggles you will be wearing will turn from a transparent to an opaque state upon movement onset, thereby blocking your vision. Therefore, in these trials you will not be able to see the target during your movement. Once you have touched the screen in the vicinity of the target, you will hear a 'beep' sound, identical to the one described earlier. Almost immediately afterwards, the goggles will turn back to their transparent state, allowing you to see how close you were to the target. In the other block, the goggles will remain transparent for the entire duration of the trial. In both conditions, please leave your finger where it landed until the target disappears.

Importantly, the touch screen sometimes fails to register touch responses. A specific angle and strength of finger press is required for the screen to register responses. In cases where the touch screen fails to register a response, wiggle your finger slightly - this should help with response detection. Please take advantage of the practice trials to familiarize yourself with the touch screen's sensitivity. Let me know if you have any questions.