A State Space Approach to Dynamic Modeling of Mouse-Tracking Data

Antonio Calcagnì[1]*, Luigi Lombardi[2], Marco D'Alessandro[2] and Francesca Freuli[2]

[1] Department of Developmental and Social Psychology, University of Padova, Padova, Italy, [2] Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

Mouse-tracking recording techniques are becoming very attractive in experimental psychology. They provide an effective means of enhancing the measurement of some real-time cognitive processes involved in categorization, decision-making, and lexical decision tasks. Mouse-tracking data are commonly analyzed using a two-step procedure which first summarizes individuals' hand trajectories with independent measures, and then applies standard statistical models on them. However, this approach can be problematic in many cases. In particular, it does not provide a direct way to capitalize the richness of hand movement variability within a consistent and unified representation. In this article we present a novel, unified framework for mouse-tracking data. Unlike standard approaches to mouse-tracking, our proposal uses stochastic state-space modeling to represent the observed trajectories in terms of both individual movement dynamics and experimental variables. The model is estimated via a Metropolis-Hastings algorithm coupled with a non-linear recursive filter. The characteristics and potentials of the proposed approach are illustrated using a lexical decision case study. The results highlighted how dynamic modeling of mouse-tracking data can considerably improve the analysis of mouse-tracking tasks and the conclusions researchers can draw from them.

Keywords: mouse tracking, state space modeling, dynamic systems, categorization task, aimed movements, Bayesian filtering

1. INTRODUCTION

Over the last decades, the study of computer-mouse trajectories has brought to light new perspectives into the investigation of a wide range of cognitive processes [e.g., for a recent review see Freeman (2017)]. Unlike traditional behavioral measures, such as reaction times and accuracies, mouse trajectories may offer a valid and cost-effective way to measure the real-time evolution of ongoing cognitive processes during experimental tasks (Friedman et al., 2013). This has also been supported by recent researches investigating mouse-tracking in association to more consolidated experimental devices, such as eye-tracking and fMRI (e.g., Quétard et al., 2016; Stolier and Freeman, 2017). In a typical mouse-tracking experiment, participants are presented with a computer-based interface showing the stimulus at the bottom of the screen and two competing categories on the left and right top corners. Participants are asked to select the most appropriate label given the task instruction and stimulus while the x-y trajectories are instantaneously recorded. The main idea is that trajectories of reaching movements can unfold the decision process underlying the hand movement behavior. For instance, the curvature of computer-mouse trajectories might reveal competing processes activated in discriminating the

two categories. Mouse-tracking has been successfully applied in several cognitive research studies, including lexical decision (Incera et al., 2017; Ke et al., 2017), social categorization (Carraro et al., 2016; Freeman et al., 2016), numerical cognition (Faulkenberry, 2014, 2016), memory (Papesh and Goldinger, 2012), moral decision (Koop, 2013), and lie detection (Monaro et al., 2017). Moreover, the availability of specialized and freely-available software for mouse-tracking experiments have strongly contributed to the wide-spread application of such a methodology in the more general psychological domain (Freeman and Ambady, 2010; Kieslich and Henninger, 2017). Recently, the debate on the nature of cognitive processes tracked by this type of reaching trajectories have also received attention from the motor control literature (Van Der Wel et al., 2009; Friedman et al., 2013).

So far, mouse-tracking data have been analyzed using simple strategies based on the conversion of x-y trajectories into summary measures, such as maximum deviation, area under the curve, response time, initiation time (Hehman et al., 2015). Although these steps are still meaningful in case of simple and well-behaved x-y trajectories, they can also provide biased results if applied to more complex and possibly noisy data. To circumvent these problems, other approaches have been proposed more recently (Cox et al., 2012; Calcagnì et al., 2017; Krpan, 2017; Zgonnikov et al., 2017). However, also the more recent proposals require modeling empirical trajectories before the data-analysis. Although these approaches potentially provide informative results in many empirical cases, they can also suffer from a number of issues, which revolve around the reduction of x-y data to simple scalar measures. For instance, problems may arise in the case of trajectories showing multiple phases, averaging with non-homogeneous curves, and signal-noise discrimination (Calcagnì et al., 2017). As far as we know, a proper framework to simultaneously model and analyse mouse-tracking data in a unified way is still lacking.

In this paper we describe an alternative perspective based on a state-space approach with the aim to simultaneously model and analyse mouse-tracking data. State-space models are very general time-series methods that allow estimating unobserved dynamics which gradually evolve over discrete time. As for diffusion models, which are widely used in modeling the temporal evolution of cognitive decision processes (Smith and Ratcliff, 2004), they belong to the general family of stochastic processes and offer optimal discrete approximation to many continuous differential systems used to represent dynamics with autoregressive patterns (Cox, 2017). In particular, we used a non-linear and discrete-time model that represents mouse trajectories as a function of some typical experimental manipulations. The model is estimated under a Bayesian framework, using a conjunction of a non-linear recursive filter and a Metropolis-Hastings algorithm. Data analyses is then performed using posterior distributions of model parameters (Gelman et al., 2014).

The reminder of this article is organized as follows. In section 2 we motivate our proposal by reviewing the main issues of a typical mouse-tracking experiment. In section 3 we present our proposal and describe its main characteristics. In section 4 we describe the application of our method to a psycholinguistic

case study. Section 5 provides a general discussion of the results, comments and suggestions for further investigations. Section 6 concludes the article by summarizing its main findings.
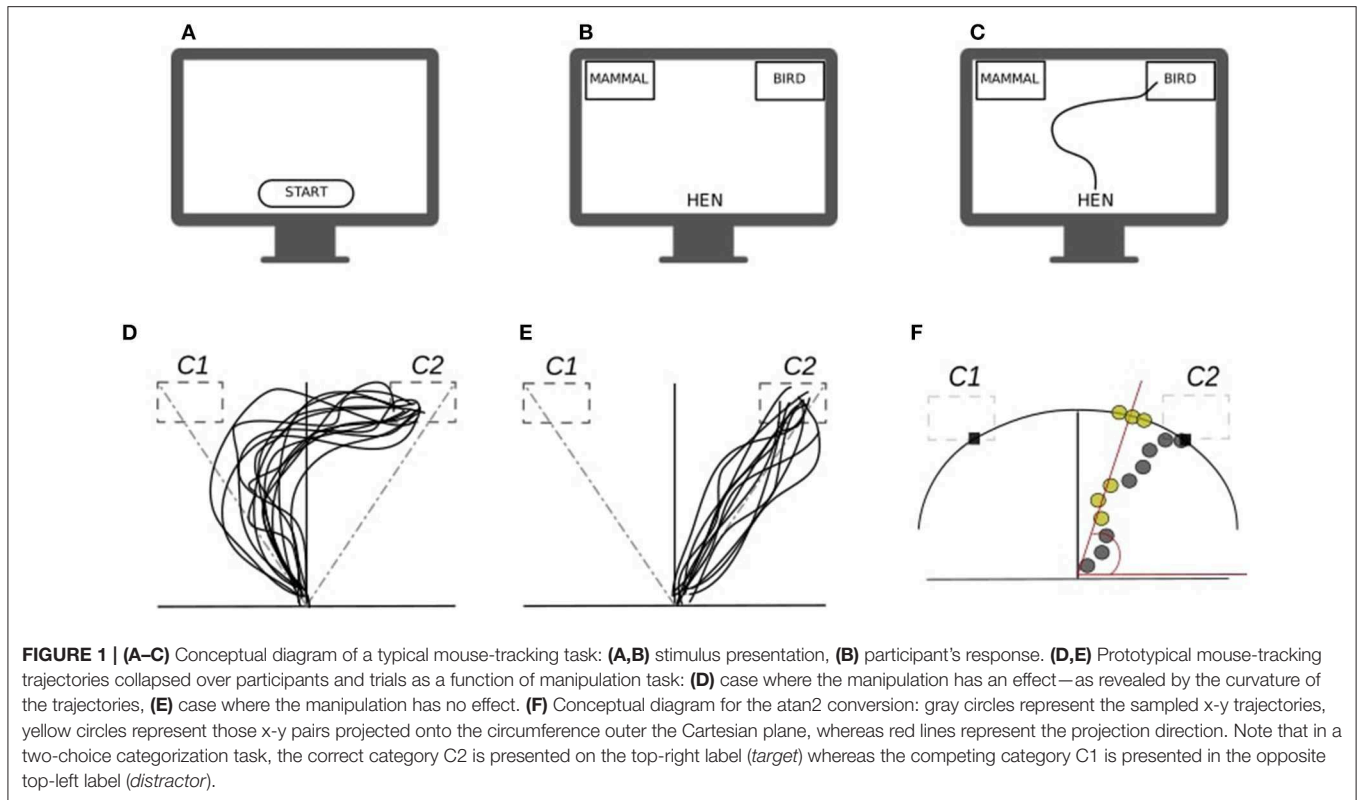
2. A MOTIVATING EXAMPLE

To begin with, consider a two-choice semantic categorization task (Dale et al., 2007), in which participants have to classify semantic stimuli (e.g., name of animals) into their corresponding categories (e.g., mammal, fish). In the most typical implementation of a mouse-tracking task, participants would sit in front of a computer screen showing a resting frame (see **Figure 1A**). They start a trial by clicking a starting button at the bottom-center of the screen, after which they are presented with a given stimulus (e.g., hen, see **Figure 1B**). To finalize the trial, participants move the cursor on the screen by means of a well-tuned computer-mouse in order to reach and select one of the two labels presented on the top-left and top-right corners of the screen (e.g., mammal vs. bird, see **Figure 1C**). In the meanwhile, x-y coordinates, initiation time, and final clicking time are registered for each participant and trial. The basic idea is that x-y trajectories reflect the extent to which the real-time categorization response is affected by the experimental manipulation. More precisely, as a result of the assumption that co-activation of competing processes continuously drive the explicit hand response (Spivey and Dale, 2006), one would suppose to see more curved—or generally irregular—trajectories in association with stimuli showing higher ambiguity. In our case, for instance, it would be expected that atypical exemplars, such as hen, dolphin, and penguin, globally produce more curved or irregular trajectories than typical exemplars like dog, rabbit, and lion (see **Figures 1D,E**).

In the mouse-tracking literature, data analysis commonly proceeds summarizing the recorded trajectories by means of few indices, which are then used as input to standard statistical techniques. In the current example, for instance, the typicality manipulation could be tested by assessing whether the distribution of maximum deviations (i.e., the maximum curvature showed by trajectories) over trials and participants is bimodal or not (Freeman and Dale, 2013). In a similar way, linear models could be employed to test whether the typicality effect varies as a function of external covariates, such as psycholinguistic variables.

However, the two-step approach does have some issues. For instance, it lacks a way to represent both the experimental variability—that is induced by task manipulations—and individual variability—that is instead produced by individual-specific motor programs. Likewise, in some cases, it might neglect relevant characteristics of x-y data, with the consequence that similar classes of trajectories are treated as if they were different. Still, a two-step approach does ignore the data generation process underlying observed trajectories. This does not allow, for example, making predictions or simulations on new data given the experimental settings.

In the next section, we will present a dynamic probabilistic model that handles mouse-tracking data in a unified way.

FIGURE 1 | (A–C) Conceptual diagram of a typical mouse-tracking task: **(A,B)** stimulus presentation, **(B)** participant's response. **(D,E)** Prototypical mouse-tracking trajectories collapsed over participants and trials as a function of manipulation task: **(D)** case where the manipulation has an effect—as revealed by the curvature of the trajectories, **(E)** case where the manipulation has no effect. **(F)** Conceptual diagram for the atan2 conversion: gray circles represent the sampled x-y trajectories, yellow circles represent those x-y pairs projected onto the circumference outer the Cartesian plane, whereas red lines represent the projection direction. Note that in a two-choice categorization task, the correct category C2 is presented on the top-right label (*target*) whereas the competing category C1 is presented in the opposite top-left label (*distractor*).

Our proposal is based upon a Bayesian non-linear state space approach, which offers a good compromise between model flexibility and model simplicity while overcoming many drawbacks of the standard mouse-tracking analyses.

3. STATE-SPACE MODELING OF MOUSE-TRACKING DATA

A state-space model is a mathematical description used to represent linear or generally non-linear dynamic models. In their general form, state-space systems consist of (i) a measurement density $f_y(\mathbf{y}_n; \mathbf{z}_n, \boldsymbol{\theta}_y)$ that describes how the observed vector of data $\mathbf{y}_n$ at time step n is linked to a possibly underlying process $\mathbf{z}_n$ and (ii) a state density $f_z(\mathbf{z}_n; \boldsymbol{\theta}_z)$ describing the transition dynamics that drive the vector of states $\mathbf{z}_n$. Temporal dynamics can be discrete or continuous and, in the latter case, stochastic differential equations are used to model the transition dynamics. By and large, there are two aims of any analysis involving state-space models. The first is to infer the unobserved process $\widetilde{\mathbf{Z}} = (\mathbf{z}_0, \ldots, \mathbf{z}_N)$ given the data $\mathbf{Y} = (\mathbf{y}_0, \ldots, \mathbf{y}_N)$. This task is usually accomplished by means of filtering and smoothing procedures (Jazwinski, 2007). The second aim regards estimating the parameters $(\boldsymbol{\theta}_y, \boldsymbol{\theta}_z)$ given the complete set of data $(\widetilde{\mathbf{Z}}, \mathbf{Y})$. This is commonly performed using gradient-based methods on the likelihood of the model (Shumway and Stoffer, 1982). Although state space models were originally used in the area of aerospace modeling (Kalman, 1960), they are now applied in a wide variety of domains, including control theory, remote

sensing, economics, and statistics (Hamilton, 1994; Shumway and Stoffer, 2006). Recently, there has also been an increasing interest in psychology, where state-space models have been used to analyse, for example, dyadic interactions (Song and Ferrer, 2009), affective dynamics (Lodewyckx et al., 2011; Bringmann et al., 2017), facial electromyography data (Yang and Chow, 2010), individual differences (Hamaker and Grasman, 2012; Chow and Zhang, 2013), and path analysis (Gu et al., 2014).

In line with this, we developed a state-space representation to simultaneously model and analyse mouse-tracking data. In particular, our proposal is to represent the empirical collection of computer-mouse trajectories as a function of two independent sub-models, one representing the experimental manipulations (*stimuli equation*) and the other capturing the main features of the mouse movement process (*states equation*). Thus, the goal of our analysis is 2-fold: (i) to determine the states equation for each participant over a set of experimental trials, (ii) to estimate the parameters governing the stimuli equation. The first goal will provide information on how participants differ from each other in terms of movement dynamics. By contrast, the second goal will find out to what extent the experimental manipulations affect the individual variations in producing mouse-tracking responses.

3.1. Data

Let $\mathbf{S}$ be a I (individuals) $\times$ J (trials) array representing the observed data. The element $\mathbf{s}_{ij}$ of $\mathbf{S}$ defines the sub-array containing the streaming of Cartesian coordinates of the computer mouse movements:

$$\mathbf{s}_{ij} = \big((\tilde{x}_0, \tilde{y}_0), \dots, (\tilde{x}_n, \tilde{y}_n), \dots, (\tilde{x}_{N_{ij}}, \tilde{y}_{N_{ij}})\big)$$

with 0 and N_{ij} being the first and the last coordinates for the i-th participant in the j-th trial. The coordinates in $\mathbf{s}_{ij}$ are temporally ordered ($0 < \dots < n < \dots < N_{ij}$) because they are usually collected while the computer-mouse is moving along its surface with a constant sampling rate. Further, to make the observed data comparable, we rescale and normalize $\mathbf{s}_{ij}$ as a function of a common ordered scale, which indicates the cumulative amount of progressive time from 0% to $N = 100\%$ (e.g., Tanawongsuwan and Bobick, 2001; Ramsay and Silverman, 2007). Thus, the final trajectories $\mathbf{s}_{ij}$ lie on the real plane defined by the hyper-rectangles $[-1, 1]^N \times [0, 1]^N$, with the first movement being equal to $(\tilde{x}_{0i}, \tilde{y}_{0i}) = (0, 0)$ by convention. Since we are interested in studying the co-activation of competing processes as reflected in some spatial properties of the response—such as *location*, *direction*, and *amplitude* of the action dynamics (Spivey and Dale, 2006; Freeman, 2017)—we need to simplify the original data structure so that these properties can easily emerge. Inspired by earlier work on this problem (Gowayyed et al., 2013; Kapsouras and Nikolaidis, 2014; Calcagnì et al., 2017), we reduce the dimensionality of the data by projecting $\mathbf{s}_{ij}$ in a proper lower-dimensional subspace of movement via the restricted four-quadrant inverse tangent mapping [atan2, see Burger and Burge (2010)] from the real coordinates to the interval $[0, \pi]^N$ as follows:

$$\underbrace{(y_0, \dots, y_n, \dots, y_N)}_{\mathbf{y}_{ij}} = \text{atan2}\big(\underbrace{(\tilde{x}_0, \tilde{y}_0), \dots, (\tilde{x}_n, \tilde{y}_n), \dots, (\tilde{x}_{N_{ij}}, \tilde{y}_{N_{ij}})}_{\mathbf{s}_{ij}}\big)$$

where y_0 is the angle at the beginning of the process whereas y_N is the angle at the end of the process. **Figure 1F** shows a graphical example of the atan2 function for a hypothetical movement path. Finally, the array of angles $\mathbf{y}_{ij}$ is the input for our state-space model.

3.2. Model Representation

The unobserved *states equation* of the model is a AR(1) Gaussian model $Z_{i,n} | Z_{i,n-1}$ with transition density equal to:

$$f(z_{i,n} | z_{i,n-1}, \theta) = \big(\sigma_i^2 \sqrt{2\pi}\big)^{-1} \cdot \exp\big(-(z_{i,n} - z_{i,n-1})^2 / 2\sigma_i^2\big) \quad (1)$$

which models how the movement process of the i-th subject changes from the step $n - 1$ to the next step n. The stochastic dynamics for the i-th subject is constrained by the variance parameter $\sigma_i^2 \in \mathbb{R}^+$ that represents the uncertainty about the future location $z_{i,n+1}$ given the current state $z_{i,n}$.

The *measurement equation* for the observations $\mathbf{y}_{ij} = (y_0, \dots, y_n, \dots, y_N)$ is modeled by means of a two-component von-Mises mixture distribution with density equal to:

$$f(y_{ijn} | \pi_{ijn}, \theta) = f(y_{ijn} | \mu_1, \kappa_1)\pi_{ijn} + f(y_{ijn} | \mu_2, \kappa_2)(1 - \pi_{ijn}) \quad (2)$$

where the generic density is the standard von-Mises law:

$$f(y_{ijn} | \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\big(\cos(y_{ijn} - \mu)^\kappa\big)$$

In the density formula, the term $I_0(.) \triangleq (2\pi)^{-1} \int_0^{2\pi} e^{\kappa \cos x} dx$ is the exponentially scaled Bessel function of order zero (Abramowitz and Stegun, 1972). The parameters of the mixture density are mapped to the experimental interface of the two-choice categorization task (see **Figures 1D,E**). In particular, the means $\{\mu_1, \mu_2\} \in [-3.14, 3.14]^2$ are mapped to the label categories *C1* and *C2* whereas the concentrations $\{\kappa_1, \kappa_2\} \in \mathbb{R}_+^2$ indicate how the observations are spread around the means. Since $\{\mu_1, \mu_2\}$ are determined by the fixed and known positions of the labels *C1* and *C2* on the screen, they are not treated as parameters to be estimated. Finally, the terms π_{ijn} and $(1 - \pi_{ijn})$ are the probabilities to activate the first and second density of the von-Mises components and are expressed as function of the latent states $\mathbf{z}_{i,0:N}$ and some additional covariates. The model is Markovian, in the sense that the unobserved states $\{Z_n; n > 1\}$ form a Markov sequence and the measurements $\{Y_n; n > 1\}$ are conditionally independent given the unobserved states.

To further characterize our state-space representation, the probability π_{ijn} is defined according to a logistic function:

$$\pi_{ijn} \triangleq \big(1 + \exp(-\beta_j - z_{i,n})\big)^{-1} \quad (3)$$

with $\beta_j \in \mathbb{R}$ being the intercept of the model. Equation (3) can be interpreted as the probability for the i-th subject at step n to categorize the j-th stimuli as belonging to *C1* (π_{ijn} tends to 1) or *C2* (π_{ijn} tends to 0). In addition, when the categories *C1* and *C2* are expressed in terms of distractor and target (Freeman, 2017), the sequences $\boldsymbol{\pi}_{ij,0:N}$ can be interpreted as the *attraction probability* that the distractor has exerted on the trajectory process $\mathbf{z}_{i,0:N}$.

The state-space representation is completed by linearly expanding the intercept term β_j as follows:

$$\beta_j \triangleq \sum_{k=1}^{K} d_{jk}\gamma_k + x_j\Big(\eta + \sum_{k=1}^{K} d_{jk}\delta_k\Big) \quad (4)$$

where $\{\gamma_k, \eta, \delta_k\} \in \mathbb{R}^3$, x_j is an element of the array $\mathbf{x} \in \mathbb{R}^J$, whereas d_{jk} is an element of the (Boolean) partition matrix $\mathbf{D}_{J \times K}$, with $d_{jk} = 1$ indicating whether the j-th stimulus belongs to the k-th level of the variable D. Note that the matrix $\mathbf{D}$ satisfies the property $\sum_{k=1}^{K} d_{jk} = 1$, for all $j = 1, \dots, J$. In our model representation, Equation (4) is the *stimuli equation* and conveys information about the experiment. It consists of three main terms. (i) A categorical term $\sum_{k=1}^{K} d_{jk}\gamma_k$ describing how the stimuli $\mathcal{J} = \{1, \dots, j, \dots, J\}$ have been arranged into $K < J$ distinct levels of a categorical variable D. (ii) A continuous term $x_j\eta$ that expresses the stimuli as a function of a continuous variable X weighted by the coefficient η. (iii) An interaction term $x_j(\eta + \sum_{k=1}^{K} d_{jk}\delta_k)$ between the levels of D and X, where $\delta_k \in \mathbb{R}$ and $\delta_1 = 0$. This definition allows for modeling all the cases implied by an univariate experimental design with at most one covariate variable. Indeed, for $\eta = 0$ and $\boldsymbol{\delta}_K = \mathbf{0}_K$ this formulation boils down to the simplest experimental case with a single categorical variable D. By contrast, for $\boldsymbol{\delta}_K = \mathbf{0}_K$ and $\boldsymbol{\gamma}_K = \mathbf{0}_K$ it reduces to the case where stimuli are simply paired with a continuous predictor X. Finally, when $\mathbf{D}_{J \times K} = \mathbf{I}_{J \times J}$,
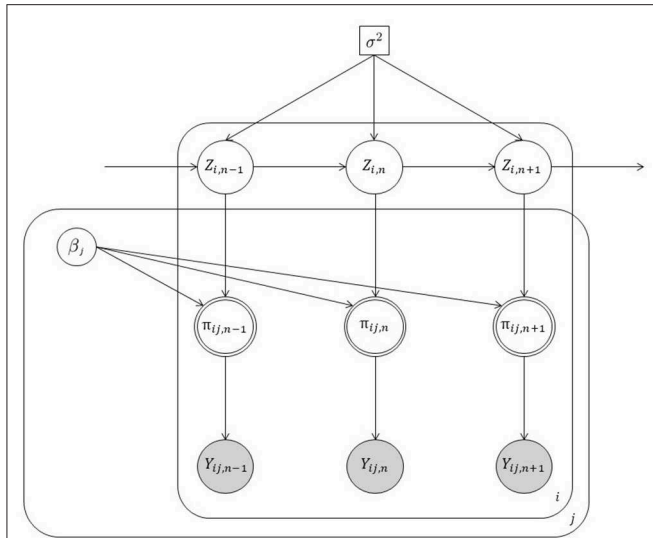
FIGURE 2 | Graphical representation of our state-space model. Note that white circles represent unobserved random variables, white double circles indicate transformed random variables, gray circles are observed random variables. Finally, square objects depict scalar quantities. Loop over individuals i, trials j, and time steps n are represented by outer squares.

the stimuli equation reduces to the most simple case where we have as many parameters as trials[1]. **Figure 2** shows a graphical representation of state-space model whereas **Figure 3** illustrates the inner-working of the model for the simplest design with a two-level experimental factor.

In our model representation, the observed movement angles $\mathbf{y}_{ij,0:N}$ are sampled from C1 (resp. C2) with probabilities equal to $\pi_{ij,0:N}$ (resp. $\pi^c_{ij,0:N} = 1 - \pi_{ij,0:N}$), which in turn are expressed as a function of the AR(1) latent trajectory $\mathbf{z}_{i,0:N}$. Hence, an increase in the latent process $z_{i,n} > 0$ will also increase the probability that y_{ijn} is sampled from the hemispace C1 (e.g., $\pi_{ijn} > 0.5$). By contrast, a decrease in the latent process $z_{i,n} < 0$ will increase the chance to sample y_{ijn} from C2 (e.g., $\pi_{ijn} < 0.5$). As a result of Equation (4) such an increasing (or decreasing) pattern can be modulated by the stimuli component $\boldsymbol{\beta}_J$. Moreover, as the coefficients $\boldsymbol{\beta}_J$ are decomposed as a function of η, $\boldsymbol{\gamma}_K$, and $\boldsymbol{\delta}_K$, we can also analyse the effect of $\boldsymbol{\beta}_J$ on $\pi_{ij,0:N}$ in terms of the experimental manipulation D, the covariate X, or the interaction term DX. **Figure 3** shows a conceptual representation of the modeling steps involved by our approach. Panel (A) shows

an example of the random-walk used to represent the movement process (Equation 3). Instead, panel (B) shows the logistic function used to form the *stimuli equation* (Equation 3) for two typical cases of $\boldsymbol{\beta}_J$. Panel (C) represents the probability $\pi_{ij,0:N}$ to activate the distractor C1 (upper panel) and the probability $\pi^c_{ij,0:N}$ to activate the target C2 (lower panel) as a function of $\mathbf{z}_{i,0:N}$ and $\boldsymbol{\beta}_J$. Finally, panel (D) depicts two cases of observed radians that are associated to $\pi_{ij,0:N}$ and $\pi^c_{ij,0:N}$. In particular, the upper panel shows an example of data with a pronounced attraction toward C1, which is in turn reflected by the blue probability curve of the panels (B,C). By contrast, the lower panel represents data with little attraction toward C1, as also reflected by the red probability curve of the panels (B,C). In this sense, as Equation (3) represents an intercept model, the parameter $\boldsymbol{\beta}_J$ does not affect the shape of the movement dynamics $\mathbf{z}_{i,0:N}$. On the contrary, it acts by shifting the movement dynamics upward ($\boldsymbol{\beta} < 0$) or downward ($\boldsymbol{\beta} > 0$) toward the C1 or C2 hemispaces, respectively.

3.3. Model Identification

State-space model identification consists of inferring the unobserved sequence of states by means of filtering and smoothing algorithms and estimating the model's parameters via Likelihood-based approximations (Shumway and Stoffer, 2006; Särkkä, 2013). For instance, in the simplest linear gaussian case, where both the states and measurement equations are linear with additive gaussian noise, inference of latent states is usually performed via Kalman filter whereas parameter's estimation is realized with the Expectation-Maximization algorithm. In our case, as Equations (3) and (4) describe a more complex non-linear model, we adopted a recursive *Gaussian approximation filter* for the inference problem (Smith and Brown, 2003), coupled with a *marginal Metropolis-Hastings* MCMC for the parameters estimation (Andrieu et al., 2010)[2].

To formulate the problem more precisely, let:

$$\boldsymbol{\Theta} = \big((\beta_1, \ldots, \beta_j, \ldots, \beta_J), (\kappa_1, \kappa_2)\big) \qquad (5)$$
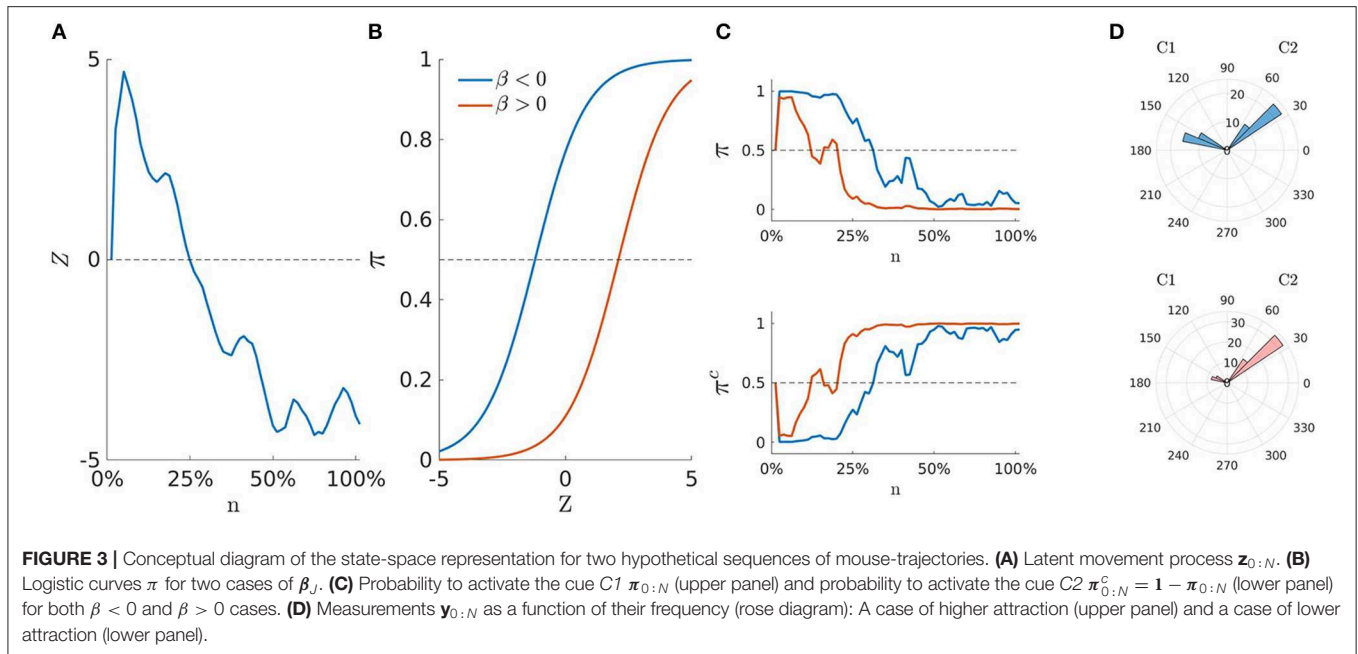
$$\mathbf{Z} = \big((z_{1,0}, \ldots, z_{1,N}), \ldots, (z_{i,0}, \ldots, z_{i,N}), \ldots, (z_{I,0}, \ldots, z_{I,N})\big) \qquad (6)$$

be the arrays representing all the $J \times 2$ unknown parameters and $I \times N$ unobserved states that characterize the model's behavior. In this context, σ_I^2 can be set to $\mathbf{1}_I$ without loss of model adequacy[3]. The joint log-density of the complete-data given the array of parameters and the observed data is defined as follows:

---

[1]To understand the meaning of the stimuli equation, consider the case of an experiment with a two-level manipulated factor A and B, each with twenty stimuli. In this case, $K = 2$ and $J = 20$ whereas $\mathbf{D}_{20 \times 2}$ is the design matrix codifying which stimulus belongs to level A ($d_{j1} = 1$, $d_{j2} = 0$) or level B ($d_{j1} = 0$, $d_{j2} = 1$). In the simple additive case, the stimuli equation is $\boldsymbol{\beta}_{20 \times 1} = \mathbf{D}_{20 \times 2} \boldsymbol{\gamma}_{2 \times 1}$ where $\boldsymbol{\gamma}$ contains the coefficients associated to the experimental levels A and B. If we also have an external covariate $\mathbf{x}$ on the stimuli, we can include this information in the stimuli equation in two ways: (i) as additive component $\boldsymbol{\beta}_{20 \times 1} = \mathbf{D}_{20 \times 2} \boldsymbol{\gamma}_{2 \times 1} + \eta \mathbf{x}_{20 \times 1}$, (ii) by including an interaction term $\boldsymbol{\beta}_{20 \times 1} = \mathbf{D}_{20 \times 2} \boldsymbol{\gamma}_{2 \times 1} + \eta \mathbf{x}_{20 \times 1} + (\mathbf{x}_{20 \times 1} \odot \mathbf{D}_{20 \times 2} \boldsymbol{\delta}_{2 \times 1})$, where $\boldsymbol{\delta}$ now codifies the interaction between the covariate and the levels A and B included in $\mathbf{D}$ (note that $\odot$ is the element-wise product). For further information on how codify categorical and continuous variables, see Fox (1997).

[2]Interestingly, this version of the MCMC algorithm can be subsumed into the more general family of particle-Metropolis Hasting (PMH) which, in turns, is a special case of Multiple Try Metropolis (MTM) techniques. For a broader review of these connections, see Martino (2018).

[3]Indeed, the constraint $\sigma_I^2 = \mathbf{1}_I$ still guarantees the mapping $\pi_{ijn} : \mathbb{R} \to [0, 1]$ to cover the needed time-to-time variability of the random walk, as Equation (3) acts as a shrinkage operator on the support of the r.vs $\{Z_{i,0}, \ldots, Z_{i,n}\}$. This has also been confirmed by several pilot simulations we ran on our model. Note that this assumption is not overly limiting, since our state-space representation is built under the *smoothness assumption* on the movement behavior of the hand, according to which large abrupt changes in the small interval $[n, n + 1]$ are not allowed (Yu et al., 2007).

FIGURE 3 | Conceptual diagram of the state-space representation for two hypothetical sequences of mouse-trajectories. **(A)** Latent movement process $\mathbf{z}_{0:N}$. **(B)** Logistic curves π for two cases of $\boldsymbol{\beta}_J$. **(C)** Probability to activate the cue $C1$ $\boldsymbol{\pi}_{0:N}$ (upper panel) and probability to activate the cue $C2$ $\boldsymbol{\pi}_{0:N}^C = \mathbf{1} - \boldsymbol{\pi}_{0:N}$ (lower panel) for both $\beta < 0$ and $\beta > 0$ cases. **(D)** Measurements $\mathbf{y}_{0:N}$ as a function of their frequency (rose diagram): A case of higher attraction (upper panel) and a case of lower attraction (lower panel).

$$\log f(\mathbf{Z}, \mathbf{Y}|\boldsymbol{\Theta}) = \log f(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\Theta}) + \log f(\mathbf{Z}|\boldsymbol{\Theta}) \qquad (7)$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \log f(\mathbf{z}_{i,0:N}|\boldsymbol{\Theta}) + \log f(\mathbf{y}_{ij,0:N}|\mathbf{z}_{i,0:N}, \boldsymbol{\Theta}) \qquad (8)$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\log f(z_{i,0}|\theta_{Z_0}) + \sum_{n=1}^{N} \log f(z_{i,n}|z_{i,n-1}, \theta_Z) \right.$$
$$\left. + \sum_{n=1}^{N} \log f(y_{ijn}|z_{i,n}, \theta_Y) \right) \qquad (9)$$

and the state and measurement equations are given as in (1) and (2) whereas the term $f(z_{ij0}|\theta_{Z_0})$ is the a-priori density function for the initial state of the process. Note that the factorization (9) is due to the Markovian properties of the model. By adopting the Bayesian perspective, we perform inference conditional on the observed sample of angles $\mathbf{Y}$, with $\boldsymbol{\Theta}$ being an unknown term. The posterior density $f(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{Y})$ of hidden states and parameters is as follows:

$$\log f(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{Y}) \propto \log f(\boldsymbol{\Theta}|\mathbf{Y}) + \log f(\mathbf{Z}|\mathbf{Y}) + \log f(\boldsymbol{\Theta}) \qquad (10)$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \log f(\boldsymbol{\Theta}|\mathbf{y}_{ij,0:N}) +$$
$$+ \sum_{i=1}^{I} \sum_{j=1}^{J} \log f(\mathbf{z}_{i,0:N}|\mathbf{y}_{ij,0:N}) + \log f(\boldsymbol{\Theta}) \qquad (11)$$

where $f(\boldsymbol{\Theta})$ is a prior density ascribed on the vector of model's parameters $\boldsymbol{\Theta}$. Note that Equation (10) comes from the standard conditional definition $f(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{Y}) = f(\mathbf{Z}, \boldsymbol{\Theta}, \mathbf{Y})/f(\mathbf{Y})$, where the joint density $f(\mathbf{Z}, \boldsymbol{\Theta}, \mathbf{Y})$ is re-arranged by factorization using

the Markovian properties of the model (e.g., see Andrieu et al., 2010). Since our aim is to get samples from the posterior $f(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{Y})$, we proceed by jointly updating $\boldsymbol{\Theta}$ and $\mathbf{Z}$ using a marginal Metropolis-Hastings. This alternates between proposing a candidate sample $\boldsymbol{\Theta}^{(t)}$ given $\boldsymbol{\Theta}^{(t-1)}$ and filtering the sequences $\mathbf{Z}^{(t)}$ conditioned on $\boldsymbol{\Theta}^{(t)}$. Finally, the candidate couple $\left(\boldsymbol{\Theta}^{(t)}, \mathbf{Z}^{(t)}\right)$ is jointly evaluated by the Metropolis-Hasting ratio.

The evaluation of both the densities $f(\mathbf{Z}|\mathbf{Y})$ and $f(\boldsymbol{\Theta}|\mathbf{Y})$ involve computing the expression in Equation (11). To do so, we derived the first term by means of filtering and smoothing procedures (Jazwinski, 2007) whereas the second term was evaluated by implementing a Metropolis-Hasting algorithm. All the technical steps for the model identification are included in **Appendices A–C** whereas all the algorithms are freely available at https://github.com/antcalcagni/SSM_mousetracking.

3.4. Model Evaluation

The state-space model formulated can be evaluated in many ways under the Bayesian framework of analysis (Shiffrin et al., 2008; Gelman et al., 2014). For instance, adequacy of the algorithm can be assessed via standard diagnostic measures, such as traceplot of the chains, autocorrelation measures, and the Gelman-Rubin statistics whereas the recovery of the true model structure can be done by simulations from the priors ascribed to the model (Gelman et al., 2014). Similarly, the adequacy of the model to reproduce the observed data can be assessed by means of simulation-based procedures (e.g., posterior predictive checks) where the fitted model is used to generate new simulated datasets that are then compared to the observed data (e.g., see Gelman et al., 1996; Cook et al., 2006). In our context, the robustness of the model formulation in recovering the true model structure as well as the goodness of fit to the observed data are assessed by

TABLE 1 | Model summary: observed and latent variables, parameters, and equations of the state-space model formulated for the analysis of mouse-tracking trajectory.

$i \in \{1, \dots, I\}$	Set index for individuals
$j \in \{1, \dots, J\}$	Set index for trials
$n \in \{0, \dots, N\}$	Set index for (discrete) time points
$\mathbf{y}_{ij} \in (0, \pi]^N$	Observed $N \times 1$ array of mouse-tracking data
$\mathbf{z}_{i,0:N} \in \mathbb{R}^N$	$N \times 1$ Array of latent states to be inferred
$\mathbf{x} \in \mathbb{R}^J$	Observed covariate of the experiment
$Y_{ijn} \sim \text{mix-vonMises}\,(\mu_1, \mu_2, \kappa_1, \kappa_2)$	Random variable governing the realization of y_{ijn}
$Z_{in}|Z_{i,n-1} \sim \text{N}(Z_{i,n-1}, \sigma_i)$	AR(1) random process governing the realization of z_{in}
$\{\mu_1, \mu_2\} \in (-\pi, \pi]^2$	Fixed parameters of the mixture von-Mises law (true locations of the stimuli)
$\{\kappa_1, \kappa_2\} \in \mathbb{R}^2$	Parameters of the von-Mises law (precision)
$\sigma_i \in \mathbb{R}_0^+$	Fixed parameter of the Gaussian law (standard deviation)
$\boldsymbol{\pi}_{ij,0:N} \in [0,1]^N$	$N \times 1$ array of attraction probability (i.e., probability to activate distractor vs. target hemi-space)
$\boldsymbol{\beta} \in \mathbb{R}^J$	$J \times 1$ array of coefficients (intercepts) modeling the experimental design
$\boldsymbol{\beta} = \mathbf{D}\boldsymbol{\gamma} + \mathbf{x}(\eta + \mathbf{D}\boldsymbol{\gamma})$	Linear equation governing the experimental design in terms of additive and interaction effects
$\mathbf{D} \in \mathbb{R}^{J \times K}$	$J \times K$ Boolean partition matrix for the experimental design
$\boldsymbol{\gamma} \in \mathbb{R}^K$	$K \times 1$ array of coefficients for the additive components of the experimental design
$(\eta, \delta) \in \mathbb{R}^{K+1}$	$(K+1) \times 1$ array of coefficients for the interaction components of the experimental design
$\boldsymbol{\Theta} = (\boldsymbol{\gamma}, \boldsymbol{\delta}, \eta, \kappa_1, \kappa_2) \in \mathbb{R}^{2K+1} \times \mathbb{R}_0^{+2}$	Complete array of parameters to be estimated (some of them can be set to zero, depending on the experimental design)

adopting a simulation-based approach. Technical details on this procedure are available in **Supplementary Materials**.

3.5. Model Summary

Table 1 shows a summary of the components of the complete state-space model used throughout the paper, including observed and latent variables, parameters and their support spaces.

4. APPLICATION

In this section, we will present an application of the model to the analysis of an already published lexical decision dataset (Barca and Pezzulo, 2012). The state-space modeling framework will be evaluated via three different instances of model representation with an increasing level of complexity. Note that the application we report here has only an illustrative purpose with the main goal to introduce and highlight the interpretation of the model's parameters and the flexibility of its representation with dynamic data. All the models were estimated using 20 (chains) $\times$ 10,000 (iterations), with a burning-in period of 2500 iterations. Starting values $\boldsymbol{\theta}_0$ for the MH algorithm were determined by maximizing the observed likelihood of the model in Equation (2). Similarly, the starting covariance matrix $\boldsymbol{\Sigma}^{(0)}$ was computed by using the Hessian of the observed likelihood at $\boldsymbol{\theta}_0$. The adaptive phase of the MH algorithm was performed at fixed interval $t + H$ (with $H = 25$) to prevent the degeneracy of the adaptation. For each model, the prior densities were defined as $f(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu} = \mathbf{0}, \mathbf{1}\sigma^2 = 25)$, where the variance was sufficiently large to cover the natural range of the model parameters. The adequacy of the model to reproduce the data was evaluated with a simulation-based approach, where a series of $M = 5,000$ new datasets $(\mathbf{Y}_1^*, \dots, \mathbf{Y}_M^*)$ were generated through the fitted model and compared with the observed data $\mathbf{Y}$ (e.g., see Cook et al., 2006). The goodness of fit was evaluated *overall* (i.e., the adequacy

of the model to reproduce the complete observed matrix $\mathbf{Y}$) and *subject-based* (i.e., the adequacy of the model to reproduce for each subject $i = 1, \dots, I$ the observed matrix $\mathbf{Y}_i$). Comparisons were computed by means of 0–100% normalized measures, with 0% indicating bad fit and 100% optimal fit. Technical details as well as extended graphical results are included in **Supplementary Materials**.

4.1. General Context and Motivation

Lexical decision is one of the most known and widely used task to study visual world recognition and reading in the cognitive psycholinguistic literature (Norris and Kinoshita, 2008; Yap et al., 2008; Hawkins et al., 2012). Generally, this task is very simple and versatile and provides an ideal context for applying the state-space modeling framework when lexical decision data are collected via the mouse tracking paradigm. In this application, we evaluated the extent to which the parameters of the state-space model reflect eventual differences associated with the manipulation of a stimulus type factor composed by words (with either high-frequency or low-frequency) and random strings (i.e., random sequence of letters that are phonotactically illegal in the language) in the lexical decision task. Moreover, we will take advantage of this psycholinguistic case study to show how our state-space model can deal with both categorical and (pseudo)quantitative predictive variables considered either individually or in interaction in the model. In particular, the first model instance will illustrate the application of our modeling framework when a simple categorical variable (stimulus type factor) is considered to affect the observed mouse-tracking trajectories collected using the lexical decision task. By contrast, the second model will be based on a simple regression-type model with a single quantitative independent variable (bigram frequency) used to predict the attraction toward the distractor category. Finally, the third model will integrate these two

TABLE 2 | Application: adequacy of the model to reproduce the observed matrices **Y** (*overall fit*) and **Y**$_i$ (*by-subject fit*).

	Overall (%)	By-subject (%)
Model 1	84	78
Model 2	73	70
Model 3	75	71

All the measures are normalized in the range 0% (bad fit)–100% (optimal fit). See **Supplementary Materials** *for technical details.*

TABLE 3 | Application: posterior means (μ), 95% posterior intervals ([$q_{0.05}, q_{0.975}$]), and Gelman-Rubin $\hat{R}$ index for the estimated parameters of Models 1–3.

		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\eta}$	$\hat{\delta}_1$	$\hat{\delta}_2$
Model 1	$q_{0.05}$	1.224	1.234	1.211			
	μ	1.323	1.337	1.310			
	$q_{0.975}$	1.443	1.457	1.432			
	$\hat{R}$	1.003	1.002	1.003			
Model 2	$q_{0.05}$				0.063		
	μ				0.078		
	$q_{0.975}$				0.091		
	$\hat{R}$				1.001		
Model 3	$q_{0.05}$	0.083	1.130	1.217	0.305	−0.505	−0.437
	μ	0.341	1.300	1.314	0.402	−0.385	−0.336
	$q_{0.975}$	0.605	1.468	1.411	0.500	−0.269	−0.235
	$\hat{R}$	1.008	1.013	1.012	1.011	1.013	1.010
All the models	$\hat{\kappa}_1 = 22.31$						
	$\hat{\kappa}_2 = 44.96$						

variables (stimulus type factor and bigram frequency) into a unified model including the main effects of the two variables as well as their interaction. In our context, the first two models will be considered as simple toy examples to illustrate the main features of the state-space model representation when applied to real data, whereas the third model will be discussed in more details according to a group analysis evaluation as well as an individual analysis representation.

4.2. Model 1
4.2.1. Data Structure and Variables
In the original work by Barca and Pezzulo (2012), the lexical decision experiment was run in Italian and based only on one stimulus type factor with four different levels: Words of high written frequency (HF, e.g., acqua "water"), words of low written frequency (LF, e.g., cervo "deer"), pseudowords (PW, e.g., "dorto"), and strings of letters that are orthographically illegal in Italian (NW, e.g., "btfpr"). In their study, participants saw a total of 96 stimuli, one at the time, and were required to categorize each stimulus as either a word or a non-word by using the mouse-tracking paradigm. Trajectories were recorded using the Mouse Tracker software (Freeman and Ambady, 2010) with sampling rate of ~70 Hz (Barca and Pezzulo, 2012). As usual, raw trajectories were normalized into $N = 101$ time steps using linear interpolation with equal spaces between coordinate samples. However, for our analysis we preferred to select only three of the four levels of the experimental factor (that is to say, HF,LF, and NW) for a total of 72 stimuli equally distributed within each level[4]. Finally, the dependent variable **Y** of Model 1 consisted of the movement angles array associated with the mouse-movement trajectory recorded for each distinct stimulus in the stimulus set.

4.2.2. Data Analysis and Results
In this first model the term β_j in the stimuli equation boils down to the simple expression:

$$\beta_j = \sum_{k=1}^{3} d_{jk}\gamma_k$$

[4]The motivation for this selection was due to some technical reasons regarding the lack of design balance in the original dataset, as the PW level showed a large number of errors when compared with the other three categories. In addition, the three-level representation of the stimulus type factor simplifies the interpretation of the results when we consider the full model with interaction.
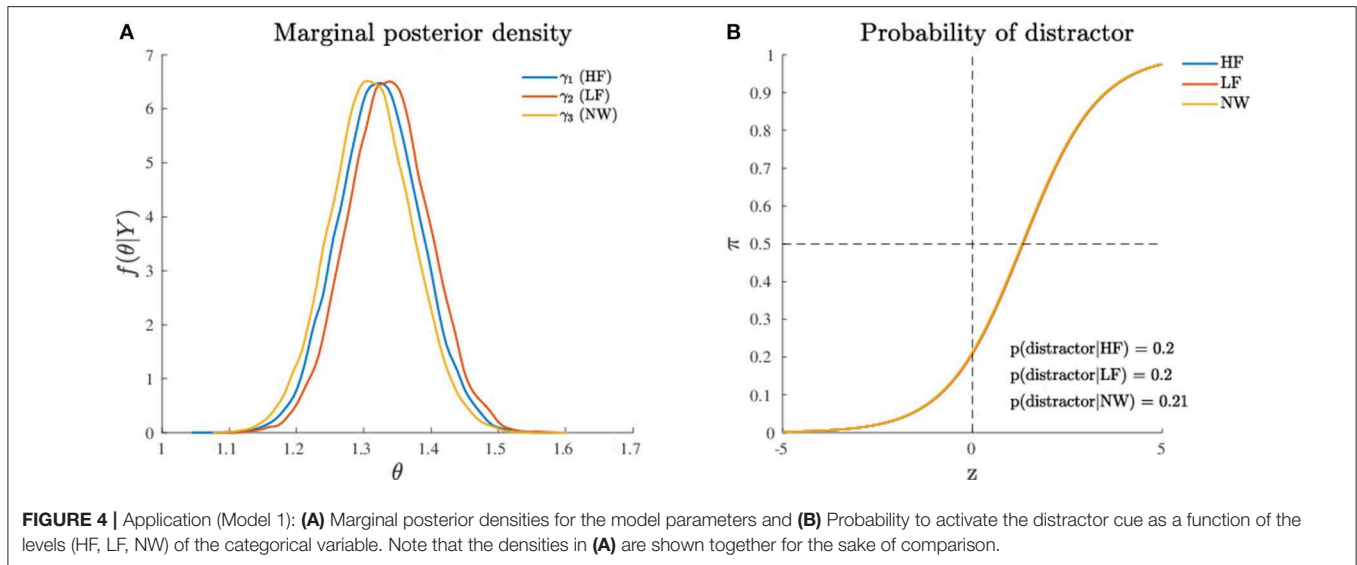
where the indices $k = 1, 2, 3$ refer to HF, LF, and NW stimuli. The MCMC convergences of the algorithm are reported in **Supplementary Materials**. The model fitted the data very satisfactorily, with overall fit of 84% and subject-based fit of 74% (see **Table 2**). The posterior quantiles (5, 50, and 95%) are reported in **Table 3** whereas **Figure 4A** shows the probability graph, that is to say, the probability to activate the distractor cue for each of the three levels HF, LF, NW as a function of the latent variable Z.

The results of this first analysis clearly show that the dynamics of the state-space model were unaffected by the different categories represented in the recoded experimental factor. This pattern finds further support in the *post-hoc* comparisons between the three experimental conditions (**Figure 4B**). In sum, these findings indicate that for a dynamic model represented according to a state-space modeling framework, the three stimulus categories (HF, LF, and NW) were all processed in a very similar way, as the original trajectories were not sufficiently different among the three stimulus categories. In substantive terms, the results of the categorical model showed how the attraction probability toward the distractor was definitively modest in all the three experimental conditions. This is evident from a direct inspection of **Figure 4B** where the probability activation function (logistic function) is shifted toward right ($Z > 0$) which in turn means that the average activation of the distractor category was relatively poor in HF, LF, and NW items. In this respect, the results of our simple spatial model were partially at odds with the outcomes observed using temporal measures (response time variables) (Barca and Pezzulo, 2012).

4.3. Model 2
4.3.1. Data Structure and Variables
Also for the second model, the dependent variable was represented by the movement angles array **Y**. However, unlike model 1, in model 2 the original independent categorical variable (stymulus type factor) was replaced with a quantitative

FIGURE 4 | Application (Model 1): **(A)** Marginal posterior densities for the model parameters and **(B)** Probability to activate the distractor cue as a function of the levels (HF, LF, NW) of the categorical variable. Note that the densities in **(A)** are shown together for the sake of comparison.

psycholinguistic variable called *bigram frequency*. Bigram frequency is defined as the frequency with which adjacent pairs of letters (bigrams) occur in printed texts; for its characteristics, it may be considered as a measure of orthographic typicality (e.g., see Hauk et al., 2006). In this second application, only bigram frequency was used as quantitative variable, since it was the only psycholinguistic variable that could be computed for all the 72 stimuli in the stimulus set. This second model instance nicely provides a simple but effective example of application of our state-space model when a continuous variable is considered to predict the attraction toward distractor.

4.3.2. Data Analysis and Results
In model 2 the term β_j simply reduces to:

$$\beta_j = x_j \eta$$

as the first and third terms in formula (4) cancel out. In this case, the variable x_j denotes the value of the bigram frequency for stimulus j in the stimulus set. For the model results, the posterior quantiles are reported in **Table 3** whereas MCMC convergences of the algorithm are reported in **Supplementary Materials**. Also in this case, the model fitted the data very well, with overall fit of 73% and subject-based fit of 70% (see **Table 2**). **Figure 5** shows the probability graph for model 2. This graph represents the probability to activate the distractor hemispace at three representative levels of the variable, i.e., the lowest, the medium, and the highest values of bigram. As evident from the graph, bigram frequency affects the probability to activate the target, with a higher probability for stimuli with low bigram frequency.

In substantive terms, the results of the quantitative model supported the evidence that the attraction probability toward the distractor was slightly affected by the specific value of the quantitative predictor (bigram frequency). In particular, low-level bigram frequencies were characterized by an average larger activation probability (0.55) for the distractor, whereas medium or large frequencies were associated with a logistic function

slightly shifted toward positive values of the latent space $Z > 0$, thus reflecting a lower chance for the distractor category (average activation probability of 0.45). Moreover, by an inspection of the contingency table for the joint representation of bigram frequency (as a transformed categorical variable) and stimulus type, we noted that low bigram frequency values were mainly characterized by string letters (NW: 94%) whereas high bigram frequency values were predominantly associated with high frequency words (HF: 55%) or low frequency words (LF: 44%).

4.4. Model 3
4.4.1. Data Structure and Variables
The final and more complex model included both the three-level categorical predictor (stimulus type factor: HF, LF, STR) and the continuous predictor (bigram frequency) as well as the interaction term between these two variables. The dependent variable was the movement angles array **Y**.

4.4.2. Data Analysis and Results
The stimuli equation which characterizes the third model is defined as follows:

$$\beta_j = \sum_{k=1}^{3} d_{jk}\gamma_k + x_j \left(\eta + \sum_{k=1}^{3} d_{jk}\delta_k \right)$$

The MCMC diagnostics together with the estimated marginal posterior densities for the model's parameters are reported in **Supplementary Materials**. The model fit was good, with an overall fit of 75% whereas the subject-based fit was equal to 71% (see **Table 2**). The posterior quantiles are reported in **Table 3**. **Figure 6** shows the probability graph for model 3. This graph represents the probability to activate the distractor hemispace for each of the three levels HF, LF, NW of the categorical factor as a function of the latent variable Z and three distinct levels (high, medium, and low) for bigram frequency. The inspection of **Figure 6** shows a clear interaction between stimulus
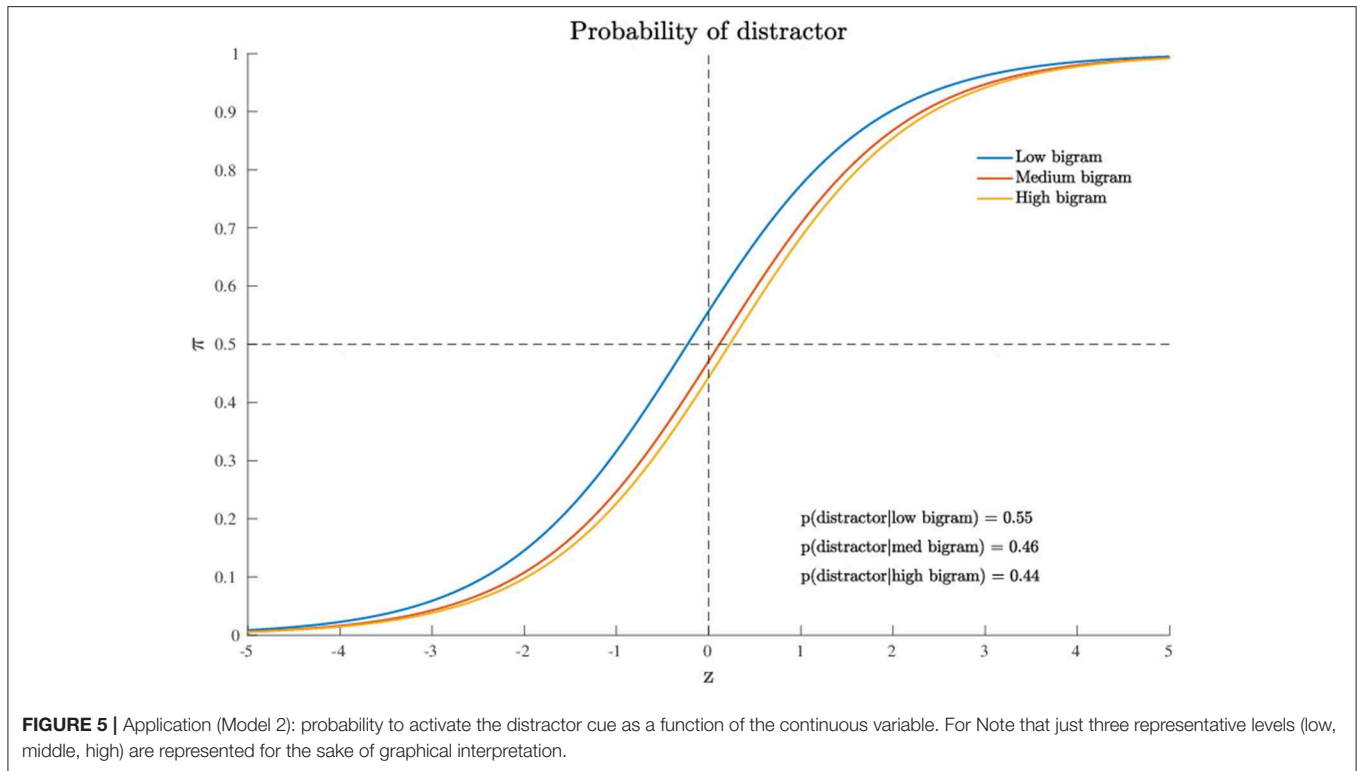
FIGURE 5 | Application (Model 2): probability to activate the distractor cue as a function of the continuous variable. For Note that just three representative levels (low, middle, high) are represented for the sake of graphical interpretation.
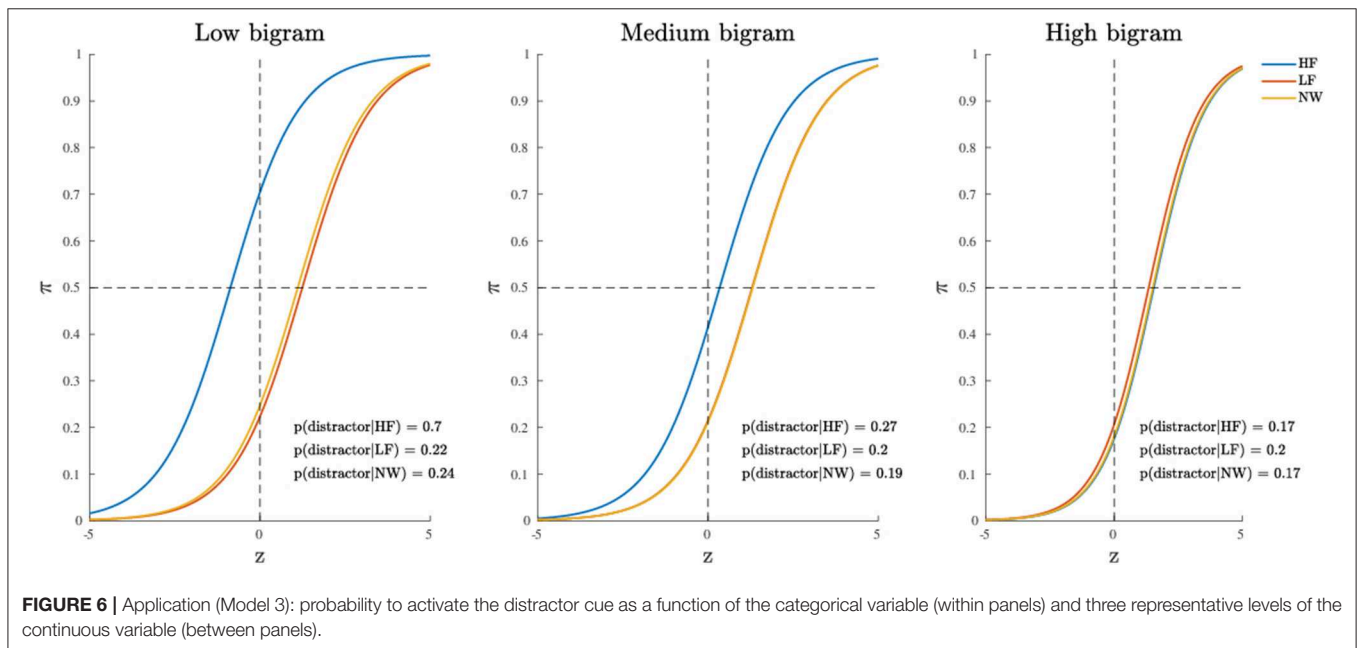

FIGURE 6 | Application (Model 3): probability to activate the distractor cue as a function of the categorical variable (within panels) and three representative levels of the continuous variable (between panels).

type factor and bigram frequency indicating that the impact of stimulus category, in particular word frequency, increases with the decrease of stimulus bigram frequency. In other words, at high level of bigram frequency, the probability to activate the distractor is similarly low in all conditions ($0.17 \leq p$-distractor ≤ 0.2). By contrast, when bigram frequency decreases—that is stimuli become orthographically atypical—the probability of

distractor activation increases, but only for the more lexically-familiar stimuli, i.e., words of high frequency (e.g., p-distractor raises from 0.17 to 0.70, in low and high bigram frequency condition, respectively).

Finally, it is also worth mentioning the emergence of the main effect of stimulus category which was instead missing in model 1. By a quick inspection of **Figure 7**, one may clearly
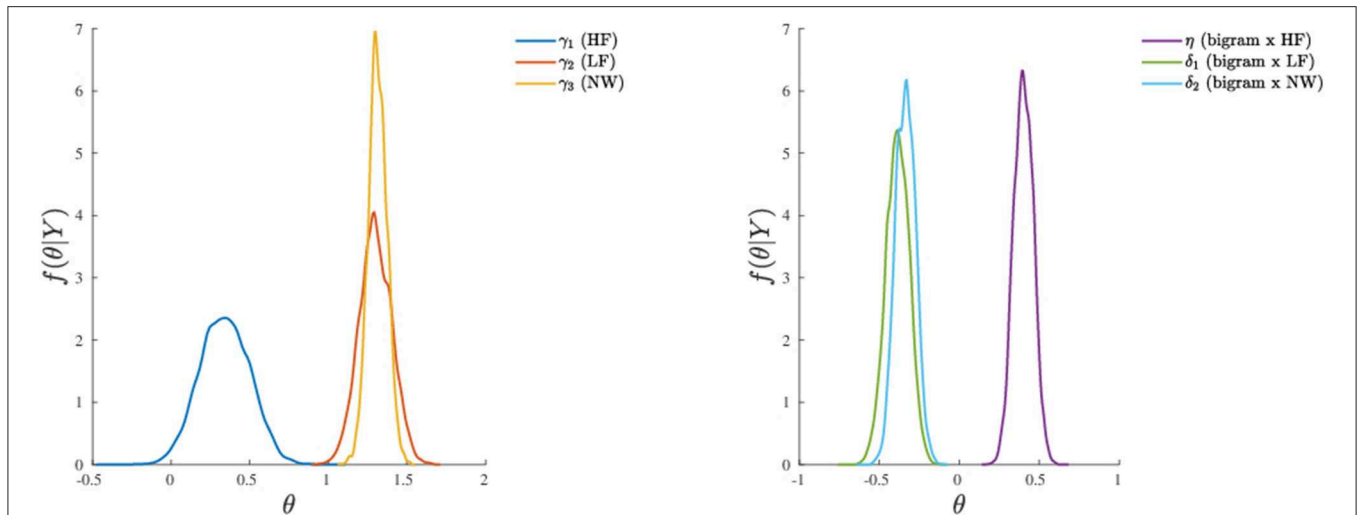
FIGURE 7 | Application (Model 3): marginal posterior densities for the model parameters. **(Left)** Parameters associated to the categorical variable. **(Right)** Parameters associated to the continuous variable and its interaction with the categorical variable. Note that the densities are shown together for the sake of comparison.

observe that HF words differ from both LF words and letter strings (NW), whereas LF words and letter strings do not differ with respect to the probability of activation of the distractor hemispace. Interestingly, the addition of the covariate bigram frequency in the model allowed the main effect of stimulus category to show up. Indeed, while at the medium and high levels of bigram frequency the results are in line with those observed at a sample level in the original study (see Figures 1, 2, 5 in the paper Barca and Pezzulo, 2012) and in a recent re-analysis (see Table 2 in the paper Calcagnì et al., 2017), in the case of low bigram the probability to activate the distractor increases with respect to high frequency words (HF). This might be somewhat related to a moderate difficulty in the orthographic processing of low frequency bigram words (e.g., see Rastle and Davis, 2008) even in the case of stimuli with richer lexical representation.

4.5. Profiles Analysis

To further investigate the dynamic characteristics involved in the lexical decision task, we extend here the results of the third model to include also a profiles analysis. **Figure 8** shows the estimated latent movement states $\mathbf{Z}_{I \times N}$ for all the participants involved in the study. The profiles appear regular, as they evolve smoothly toward the target cue (T). We grouped the dynamics into four well-separated clusters (**Figure 8**, smallest panels on the right) according to their functional similarities (Ramsay and Silverman, 2007). Particularly, the first group is characterized by a higher exploration of the distractor's hemispace, especially in the first 30% of the process. The same applies to the third and fourth groups, although they show a gradual activation of the distractor. Finally, the second group clearly represents those profiles with no uncertainty in the categorization process, as they show no activation of the distractor's hemispace at all. Although well-separated among them, these clusters still show some level of inner heterogeneity (for example, see group 1 and 4). To study this latter issue in terms of experimental manipulations, we

focused on group 1 and considered the low vs. high frequency conditions (HF vs. LF). We also selected the middle phase of the process ($\Delta = 30 - 50\%$), where it is expected to observe larger cognitive competitions in the categorization (Barca and Pezzulo, 2012). **Figure 9** shows the participants' profiles in terms of attraction probability $\boldsymbol{\pi}_{4 \times N}$ for the two lexical conditions. As expected, the profiles differ between these conditions, with LF eliciting higher attraction probability. This is in line with the fact that low frequency words have a weaker lexical representation than high frequency stimuli and consequently they are more difficult to process (Barca and Pezzulo, 2012). Interestingly, the individual profiles also differ in the way they activate the distractor. For instance, the participant 6 had higher probability in both LF ($p_\Delta(D) = 0.67$) and HF ($p_\Delta(D) = 0.54$) conditions whereas the participant 7 had a more pronounced activation just in the LF condition ($p_\Delta(D) = 0.57$) than HF ($p_\Delta(D) = 0.43$). Similarly, participants 6 and 7 seemed to prolong the competing dynamics up to the 50% of the process, by contrast participants 8 and 15 seemed to resolve the lexical competition earlier as showed by the abrupt decreasing of their curves. We complete our analysis by evaluating how individual profiles are linked to empirical measurements. **Figure 10** represents this scenario for two stimuli belonging to HF and LF conditions. As we can notice, the curves present the same dynamics (due to the latent states $\mathbf{z}_{i,0:N}$) although they clearly differ in terms of attraction exerted by the stimulus (due to the β component of the model). In this case, the LF stimulus produced larger conflict than HF in the lexical categorization. This is evident when we turn back to the observed data: as expected, the rose diagrams of LF showed larger directions in the distractor's hemispace.

5. DISCUSSION

We have described a new approach to model and analyse dynamic data coming from mouse-tracking experiments. Our
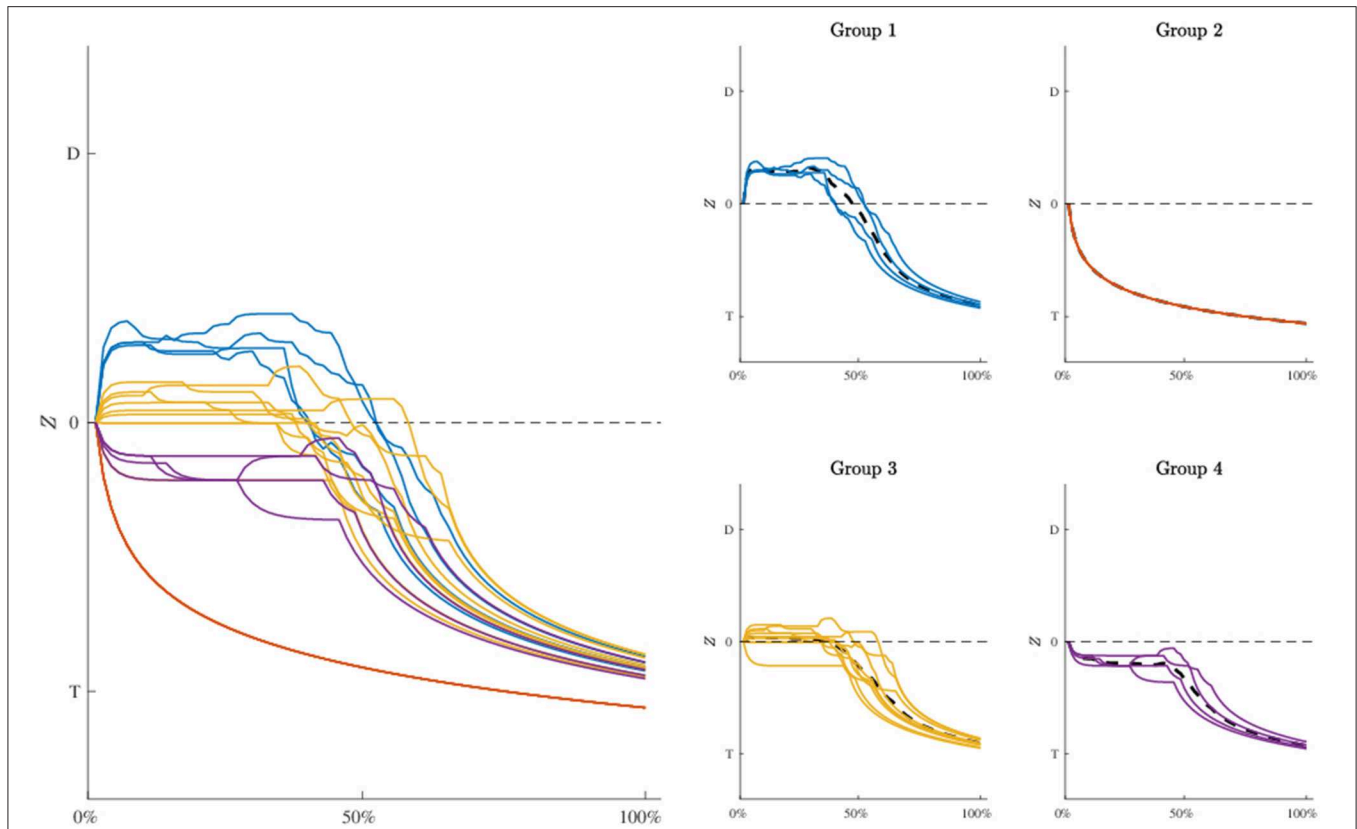
FIGURE 8 | Application: estimated movement dynamics $\mathbf{z}_{i,0:N}$ of each participant (biggest panel, **Left**) and clusters of profiles in terms of their functional similarity (smallest panels, **Right**). Note that averaged profiles are represented as dashed lines whereas D and T in all the panels indicate distractor and target, respectively. Groups' composition: participants 6, 7, 8, 15 (group 1), 1, 4, 19, 21 (group 2), 2, 3, 5, 12, 13, 16, 17, 20, 22 (group 3), 10, 11, 14, 18 (group 4).
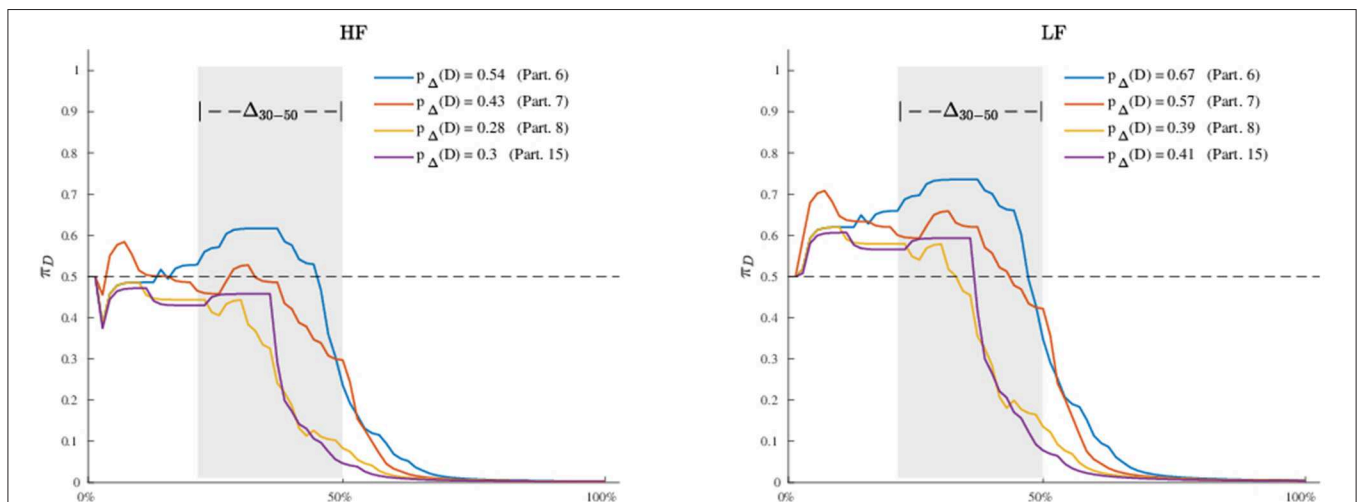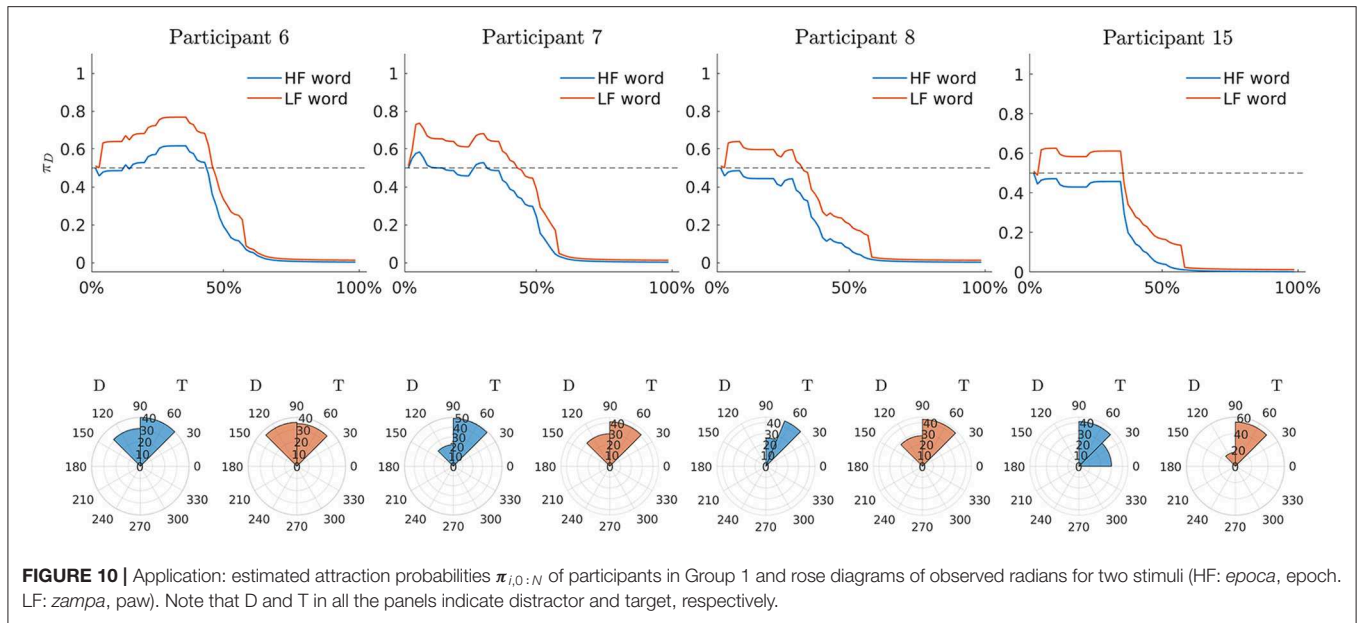

FIGURE 9 | Application: estimated attraction probabilities $\boldsymbol{\pi}_{i,0:N}$ of participants in Group 1 for the high frequency **(Left)** and low frequency **(Right)** lexical conditions. Note that the probability curves are computed with respect to the distractor (D), the gray area in both panels indicates a selected window of processing ($\Delta = 30 - 50\%$), whereas the terms $p_\Delta(D)$ are computed using a normalized discrete approximation of the integral of the probability curves in the selected process window Δ.

proposal took the advantages of a state-space representation, in which the observed data $\mathbf{Y}$ were thought as being function of two independent sub-models, one representing the movement process and its properties ($\mathbf{Z}$) and the second modeling the two-choice experimental task ($\boldsymbol{\beta}$) according to which the data were collected. These sub-models were integrated by means of an

FIGURE 10 | Application: estimated attraction probabilities $\boldsymbol{\pi}_{i,0:N}$ of participants in Group 1 and rose diagrams of observed radians for two stimuli (HF: *epoca*, epoch. LF: *zampa*, paw). Note that D and T in all the panels indicate distractor and target, respectively.

inverse-logit function (π) that expressed how the experimental manipulations acted on the movement processes in selecting the final correct response against the competing one. This formulation was flexible enough to take into account the complexity of some dynamic behaviors showed by the reaching trajectories. Moreover, it allowed for separately accounting for the motor heterogeneity and experimental variability in **Y**. Indeed, when $\boldsymbol{\beta} = \mathbf{1}\beta_0$ our state-space representation simply reduced to a model where the experimental manipulations had no relevant effect in reproducing the observed data. This instance has been illustrated in section 4.3 (Model 1). In this case, as **Z** = **0** was not allowed in our model formulation, all the variability of **Y** can be ascribed to **Z**. This is relevant in view of the fact that movement variability may reflect only individual motor executions in absence of any experimental manipulations (Yu et al., 2007). The movement component **Z** was modeled to be Markovian with gaussian transition density.

Although more complex models can be used to represent movement dynamics, simple random walks still allows a great deal of flexibility in modeling reaching trajectories under weak assumptions on the movement behavior (e.g., see Yu et al., 2007; Paninski et al., 2010). In particular, in the case of mouse-tracking tasks, they allow representations of the following three properties: (i) Each movement is goal-oriented as individuals have to finalize the action by clicking on one of the two categories shown on the screen. (ii) Mouse-tracking trajectories generally start at rest, proceed out in the movement space, and end at rest. (iii) Hand trajectories tend to be smooth during the reaching process, i.e., small changes in the interval $[n, n + 1]$ are more likely than large and abrupt changes (Brockwell et al., 2004; Spivey et al., 2010). The stimuli component $\boldsymbol{\beta}$ was defined to be a linear combination of information typically involved in a univariate design, namely a categorical variable D containing the levels of the experimental factor and a continuous covariate

X. This gave researchers the opportunity to additionally analyse which component of the experimental design is relevant in producing the effect of $\boldsymbol{\beta}$ on **Y**. The data-generation process was defined according to a mixture of two von-Mises distributions representing the categories of a two-choice categorization task. Among others, we chose this distribution because it provides a flexible representation for angular ordered data, especially because it simplifies mathematical computations involved in the model's derivation (e.g., see McClintock et al., 2012; Mulder and Klugkist, 2017).

There are other existing methods that offer alternative ways to model mouse-tracking data. For instance, Van Der Wel et al. (2009) proposed the use of the *movement superposition model* (Henis and Flash, 1995) to model and analyse the typical two-choice lexical decision task. In particular, they modeled mouse-tracking trajectories by representing the complete hand movement as a summation of sub-movements, which were obtained by the solution of the minimum-jerk equation for the standard reaching trajectory (i.e., a movement characterized by a bell-shaped speed profile that minimizes the sum of the squared rates of jerks over the movement duration). Similarly, Friedman et al. (2013) discussed how an intermittent model of arms movement can be used for reaching trajectories in random-dot experiments. They used both Wiener's diffusion process and Flash and Hogan's movement equation to predict reaction times (RTs) and movement data. Their goal was to assess the link between movement trajectories and underlying cognitive processing. Our model differs in some respects from these works. With regards to Van Der Wel et al. (2009), for instance, we used a stochastic state-space approach to model the movement trajectories instead of deterministic equations. Instead, with respect to Friedman et al. (2013), we tailor-made our model to a typical two-choice categorization task, making use of few assumptions on the nature of the movement process

[as those implied by the Gaussian AR(1) process]. By and large, our goal was not to provide a mathematical representation of the cognitive components underpinning mouse-trajectories since the model does not describe the cognitive processing *per se*. By contrast, we simply provided a *statistical model* for the analysis of mouse-tracking data, which can offer a good compromise between data modeling and data analysis.

5.1. Model's Advantages and Limitations

Our non-linear state-space model has several advantages. For instance, when comparing with the standard approaches, our proposal provides a unified analytic framework to simultaneously model and analyse trajectories data. By modeling movement heterogeneity and task variability together, we can evaluate how experimental variables directly act on the observed series of trajectories, with no need to use any kind of summary measures. An additional advantage of our model concerns the study of individual differences in terms of latent dynamics. While this is impractical in standard two-step approaches, in our proposal researchers can assess individual variations by studying the movement profiles $\tilde{\mathbf{Z}}$ once they are estimated. For instance, they can be analyzed in terms of similarity/dissimilarity with regards to external individual covariates (e.g., vocabulary knowledge and bilingualism in psycholinguistic experiments; IQ, risk-taking propensity, or more generally clinical variables in decision-making tasks). Still, individual dynamics can be compared each other qualitatively in terms of chance to activate the distractor or target cues. As the dynamics are normalized on a common cumulative scale, researchers can assess whether the chance to activate the distractor cue at a certain percentage of the process and for an experimental manipulation, is particularly higher in a sub-group of participants (this case, for example, has been described in section 4.6).

As for any modeling approach, also the current proposal can potentially suffer from some limitations. A first limitation concerns the only-intercept model $\pi(\mathbf{Z}, \boldsymbol{\beta})$ used to integrate individual dynamics and experimental information. Although this was enough to represent whether or not certain stimuli can increase the probability to select the distractor cue, we may want to known whether some experimental manipulations can modify the individual dynamics as well. However, this would particularly pronounce the computational costs required for the model identification (especially with regards to filtering), as we need to appropriately generalize Equation (3) to include more parameters. Lastly, in the current study we used univariate non-linear state-space models to represent individual dynamics for the sake of parsimony. However, more complicated situations may require models including further movement characteristics like step-length, velocity, acceleration, and jerk (Kulkarni and Paninski, 2008), which may be modeled as statistical constraints of the model (Ciavolino and Calcagnì, 2014; Calcagnì et al., 2017).

5.2. Further Extensions

Our non-linear state-space model can be improved in many aspects. For instance, the stimuli equation (4) can be generalized to cope with more complex experimental designs, like those involving multiple factors and covariates together with high-order interaction terms. Likewise, the current model restrictions can be relaxed to allow changes in slopes of $\pi(\mathbf{Z}, \boldsymbol{\beta})$ as a function of the experimental stimuli. Further, the development of a hierarchical representation of the model, with a random-effect component in the state Equation (3), would offer a way to model the inter-individual variations as resulting from an underlying common population. Still, the development of a multivariate state-space model to include other movement components will surely be considered a future extension of the present work. Further studies may lead to generalize the AR(1) process used for the movement dynamics to include former knowledge on the deterministic constraints of the hand movement as those used, for instance, by Van Der Wel et al. (2009) and Friedman et al. (2013). Moreover, further studies may also lead to generalize the AR(1) process used for the movement dynamics to include former knowledge on the deterministic constraints of the hand movement as those used, for instance, by Van Der Wel et al. (2009) and Friedman et al. (2013). Finally, an open issue which deserves greater consideration in future investigations is the need for a formal comparative framework with which we may eventually contrast and compare spatial modeling perspectives (like the one presented in this contribution) and currently used methods for analyzing mouse tracking data based on descriptive statistics (e.g., see Freeman, 2017).

6. CONCLUSIONS

In this paper we presented a novel and comprehensive analytic framework for modeling and analyse mouse-tracking trajectories. In particular, a non-linear state-space approach was used to model the observed trajectories as a function of both individual movement dynamics and experimental variables. Model identification was performed under the umbrella of Bayesian methods, in which a Metropolis-Hastings algorithm was coupled with a recursive gaussian approximation filter to get posterior distributions of model parameters. For the sake of illustration, we applied our new approach to a real mouse-tracking dataset concerning a two-choice lexical categorization task. The results indicated how our proposal can provide valuable insights to assess the dynamics involved in the decision task and identify how the experimental variables significantly contributed to the observed movement heterogeneity. Moreover, the analysis of individual profiles allowed for comprehensive and reliable identification of individual and group-based differences in the dynamics of decision making.

In conclusion, we think that this work yielded interesting findings in the development of computational models able to capture the unfolding high-level cognitive processes as reflected by motor executions which are typically involved in mouse-tracking tasks. To our knowledge, this is the first time that mouse-tracking data are fully modeled and analyzed within a process-oriented approach. We believe our contribution will offer a novel strategy that may help cognitive researchers to understand the roles of cognition and action in mouse-tracking based experiments.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this manuscript are not publicly available. Requests to access the datasets should be directed to Laura Barca, laurabarcapst@gmail.com.

AUTHOR CONTRIBUTIONS

AC and LL contributed to the conceptualization and statistical modelization. AC, MD'A, and LL performed the main statistical analyses. AC and MD'A performed the supplementary statistical analyses. FF contributed to the statistical analyses and revised supplementary statistical analyses. MD'A contributed to the section Model Evaluation. FF contributed to the section Application. AC wrote the first draft of the study. AC, LL, MD'A, and FF revised the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02716/full#supplementary-material

REFERENCES

Abramowitz, M., and Stegun, I. A. (1972). *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. New York, NY: Dover.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *J. R. Stat. Soc. B Stat. Methodol.* 72, 269–342. doi: 10.1111/j.1467-9868.2009.00736.x

Andrieu, C., and Roberts, G. O. (2009). The pseudo-marginal approach for efficient monte carlo computations. *Ann. Stat.* 37, 697–725. doi: 10.1214/07-AOS574

Ansley, C. F., and Kohn, R. (1982). A geometrical derivation of the fixed interval smoothing algorithm. *Biometrika* 69, 486–487.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von mises-fisher distributions. *J. Mach. Learn. Res.* 6, 1345–1382.

Barca, L., and Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE* 7:e35932. doi: 10.1371/journal.pone.0035932

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., and Tuerlinckx, F. (2017). Changing dynamics: time-varying autoregressive models using generalized additive modeling. *Psychol. Methods* 22, 409–425. doi: 10.1037/met0000085

Brockwell, A. E., Rojas, A. L., and Kass, R. (2004). Recursive bayesian decoding of motor cortical signals by particle filtering. *J. Neurophysiol.* 91, 1899–1907. doi: 10.1152/jn.00438.2003

Burger, W., and Burge, M. J. (2010). *Principles of Digital Image Processing: Fundamental Techniques*. London: Springer Science & Business Media.

Cabras, S., Nueda, M. E. C., and Ruli, E. (2015). Approximate bayesian computation by modelling summary statistics in a quasi-likelihood framework. *Bayesian Anal.* 10, 411–439. doi: 10.1214/14-BA921

Calcagnì, A., Lombardi, L., and Sulpizio, S. (2017). Analyzing spatial data from mouse tracker methodology: an entropic approach. *Behav. Res. Methods* 49, 2012–2030. doi: 10.3758/s13428-016-0839-5

Carraro, L., Castelli, L., and Negri, P. (2016). The hand in motion of liberals and conservatives reveals the differential processing of positive and negative information. *Acta Psychol.* 168, 78–84. doi: 10.1016/j.actpsy.2016.04.006

Chow, S.-M., and Zhang, G. (2013). Nonlinear regime-switching state-space (RSSS) models. *Psychometrika* 78, 740–768. doi: 10.1007/s11336-013-9330-8

Ciavolino, E., and Calcagnì, A. (2014). A generalized maximum entropy (GME) approach for crisp-input/fuzzy-output regression model. *Qual. Quant.* 48, 3401–3414. doi: 10.1007/s11135-013-9963-9

Cook, S. R., Gelman, A., and Rubin, D. B. (2006). Validation of software for Bayesian models using posterior quantiles. *J. Comput. Graph. Stat.* 15, 675–692. doi: 10.1198/106186006X136976

Cox, D. R. (2017). *The Theory of Stochastic Processes*. New York, NY: Routledge.

Cox, G., Kachergis, G., and Shiffrin, R. (2012). "Gaussian process regression for trajectory analysis," in *Proceedings of the Cognitive Science Society*, Austin, TX: Cognitive Science Society, Vol. 34.

Dale, R., Kehoe, C., and Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Mem. Cogn.* 35, 15–28. doi: 10.3758/BF03195938

De Jong, P. (1988). The likelihood for a state space model. *Biometrika* 75, 165–169.

Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Can. J. Exp. Psychol.* 68, 147–151. doi: 10.1037/cep0000021

Faulkenberry, T. J. (2016). Testing a direct mapping versus competition account of response dynamics in number comparison. *J. Cogn. Psychol.* 28, 825–842. doi: 10.1080/20445911.2016.1191504

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications, Inc.

Freeman, J. B. (2017). Doing psychological science by hand. *Curr. Dir. Psychol. Sci.* 27, 315–323. doi: 10.1177/0963721417746793

Freeman, J. B., and Ambady, N. (2010). Mousetracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* 42, 226–241. doi: 10.3758/BRM.42.1.226

Freeman, J. B., and Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behav. Res. Methods* 45, 83–97. doi: 10.3758/s13428-012-0225-x

Freeman, J. B., Pauker, K., and Sanchez, D. T. (2016). A perceptual pathway to bias: interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychol. Sci.* 27, 502–517. doi: 10.1177/0956797615627418

Friedman, J., Brown, S., and Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *J. Math. Psychol.* 57, 140–151. doi: 10.1016/j.jmp.2013.06.005

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*, Vol. 2. Boca Raton, FL: CRC Press.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sinica* 6, 733–760.

Gowayyed, M. A., Torki, M., Hussein, M. E., and El-Saban, M. (2013). "Histogram of oriented displacements (HOD): describing trajectories of human joints for action recognition," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (AAAI Press), 1351–1357.

Gu, F., Preacher, K. J., and Ferrer, E. (2014). A state space modeling approach to mediation analysis. *J. Educ. Behav. Stat.* 39, 117–143. doi: 10.3102/1076998614524823

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli* 7, 223–242. doi: 10.2307/3318737

Hamaker, E., and Grasman, R. (2012). Regime switching state-space models applied to psychological processes: handling missing data and making inferences. *Psychometrika* 77, 400–422. doi: 10.1007/s11336-012-9254-8

Hamilton, J. D. (1994). State-space models. *Handb. Econometr.* 4, 3039–3080.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., and Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed

by linear regression analysis of ERP data. *Neuroimage* 30, 1383–1400. doi: 10.1016/j.neuroimage.2005.11.048

Hawkins, G., Brown, S. D., Steyvers, M., and Wagenmakers, E.-J. (2012). Decision speed induces context effects in choice. *Exp. Psychol.* 59, 206–215. doi: 10.1027/1618-3169/a000145

Hehman, E., Stolier, R. M., and Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Process. Intergr. Relat.* 18, 384–401. doi: 10.1177/1368430214538325

Henis, E. A., and Flash, T. (1995). Mechanisms underlying the generation of averaged modified trajectories. *Biol. Cybernet.* 72, 407–419.

Incera, S., Shah, A. P., McLennan, C. T., and Wetzel, M. T. (2017). Sentence context influences the subjective perception of foreign accents. *Acta Psychol.* 172, 71–76. doi: 10.1016/j.actpsy.2016.11.011

Jazwinski, A. H. (2007). *Stochastic Processes and Filtering Theory*. New York, NY: Courier Corporation.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45.

Kapsouras, I., and Nikolaidis, N. (2014). Action recognition on motion capture data using a dynemes and forward differences representation. *J. Visual Commun. Image Represent.* 25, 1432–1445. doi: 10.1016/j.jvcir.2014.04.007

Ke, A. H., Epstein, S. D., Lewis, R., and Pires, A. (2017). The quantificational domain of DOU: an experimental study. *J. Psycholinguist. Res.* 47, 537–556. doi: 10.1007/s10936-017-9532-9

Kieslich, P. J., and Henninger, F. (2017). Mousetrap: an integrated, open-source mouse-tracking package. *Behav. Res. Methods* 49, 1652–1667. doi: 10.3758/s13428-017-0900-z

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judg. Decis. Making* 8, 527–539.

Krpan, D. (2017). Behavioral priming 2.0: enter a dynamical systems perspective. *Front. Psychol.* 8:1204. doi: 10.3389/fpsyg.2017.01204

Kulkarni, J. E., and Paninski, L. (2008). State-space decoding of goal-directed movements. *IEEE Signal Process. Mag.* 25, 78–86. doi: 10.1109/MSP.2008.4408444

Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N., and Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *J. Math. Psychol.* 55, 68–83. doi: 10.1016/j.jmp.2010.08.004

Luengo, D., and Martino, L. (2013). "Fully adaptive gaussian mixture metropolis-hastings algorithm," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE), 6148–6152.

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handb. Stat.* 25, 459–507. doi: 10.1016/S0169-7161(05)25016-2

Martino, L. (2018). A review of multiple try mcmc algorithms for signal processing. *Digital Signal Process.* 75, 134–152. doi: 10.1016/j.dsp.2018.01.004

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecol. Monogr.* 82, 335–349. doi: 10.1890/11-0326.1

Mendel, J. M. (1995). *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Upper Saddle River, NJ: Pearson Education.

Monaro, M., Gamberini, L., and Sartori, G. (2017). The detection of faked identity using unexpected questions and mouse dynamics. *PLoS ONE* 12:e0177851. doi: 10.1371/journal.pone.0177851

Mulder, K., and Klugkist, I. (2017). Bayesian estimation and hypothesis tests for a circular generalized linear model. *J. Math. Psychol.* 80, 4–14. doi: 10.1016/j.jmp.2017.07.001

Norris, D., and Kinoshita, S. (2008). Perception as evidence accumulation and bayesian inference: insights from masked priming. *J. Exp. Psychol. Gen.* 137, 434–455. doi: 10.1037/a0012799

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rahnama Rad, K., Vidne, M., et al. (2010). A new look at state-space models for neural data. *J. Comput. Neurosci.* 29, 107–126. doi: 10.1007/s10827-009-0179-x

Papesh, M. H., and Goldinger, S. D. (2012). Memory in motion: movement dynamics reveal memory strength. *Psychonom. Bull. Rev.* 19, 906–913. doi: 10.3758/s13423-012-0281-3

Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., et al. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: evidence from eye and mouse tracking in visual search. *J. Vis.* 16, 28–28. doi: 10.1167/16.11.28

Ramsay, J. O., and Silverman, B. W. (2007). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

Rastle, K., and Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Lang. Cogn. Process.* 23, 942–971. doi: 10.1080/01690960802069730

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*, Vol. 3. New York, NY: Cambridge University Press.

Shampine, L. F. (2008). Vectorized adaptive quadrature in matlab. *J. Comput. Appl. Math.* 211, 131–140. doi: 10.1016/j.cam.2006.11.021

Shiffrin, R. M., Lee, M. D., Kim, W., and Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cogn. Sci.* 32, 1248–1284. doi: 10.1080/03640210802414826

Shumway, R. H., and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the em algorithm. *J. Time Series Anal.* 3, 253–264.

Shumway, R. H., and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications: With R Examples*. Springer Science & Business Media.

Smith, A. C., and Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Comput.* 15, 965–991. doi: 10.1162/089976603765202622

Smith, P. L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci.* 27, 161–168. doi: 10.1016/j.tins.2004.01.006

Song, H., and Ferrer, E. (2009). State-space modeling of dynamic psychological processes via the kalman smoother algorithm: rationale, finite sample properties, and applications. *Struct. Equat. Model.* 16, 338–363. doi: 10.1080/10705510902751432

Spivey, M. J., and Dale, R. (2006). Continuous dynamics in real-time cognition. *Curr. Dir. Psychol. Sci.* 15, 207–211. doi: 10.1111/j.1467-8721.2006.00437.x

Spivey, M. J., Dale, R., Knoblich, G., and Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der wel et al. (2009). *J. Exp. Psychol.* 36, 251–254. doi: 10.1037/a0017170

Stolier, R. M., and Freeman, J. B. (2017). A neural mechanism of social categorization. *J. Neurosci.* 37, 5711–5721. doi: 10.1523/JNEUROSCI.3334-16.2017

Tanawongsuwan, R., and Bobick, A. (2001). "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001*, Vol. 2.

Tanner, M. A. (1991). *Tools for Statistical Inference*, Vol. 3. New York, NY: Springer.

Van Der Wel, R. P., Eder, J. R., Mitchel, A. D., Walsh, M. M., and Rosenbaum, D. A. (2009). Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: a commentary on Spivey, Grosjean, and Knoblich (2005). *J. Exp. Psychol.* 35, 588–594. doi: 10.1037/0096-1523.35.2.588

Yang, M., and Chow, S.-M. (2010). Using state-space model with regime switching to represent the dynamics of facial electromyography (EMG) data. *Psychometrika* 75, 744–771. doi: 10.1007/s11336-010-9176-2

Yap, M. J., Balota, D. A., Tse, C.-S., and Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: evidence for opposing interactive influences revealed by RT distributional analyses. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 495–513. doi: 10.1037/0278-7393.34.3.495

Yu, B. M., Kemere, C., Santhanam, G., Afshar, A., Ryu, S. I., Meng, T. H., et al. (2007). Mixture of trajectory models for neural decoding of goal-directed movements. *J. Neurophysiol.* 97, 3763–3780. doi: 10.1152/jn.00482.2006

Zgonnikov, A., Aleni, A., Piiroinen, P., O'Hora, D., and di Bernardo, M. (2017). Decision landscapes: visualizing mouse-tracking data. *Open Sci.* 4:170482. doi: 10.1098/rsos.170482

Appendix A: Filtering and smoothing

The term $f(\mathbf{Z}|\mathbf{Y})$ in Eq. (11) is recursively computed given all the measurements up to the n-th step. Let:

$$\log f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta) \propto \log f(y_{ijn}|z_{i,n},\theta)+ \tag{A.1}$$
$$+ \log \int_{\mathbb{R}} f(z_{i,n}|z_{i,n-1},\theta)f(z_{i,n-1}|\mathbf{y}_{ij,0:n-1},\theta)dz_{i,n-1}$$

be the filtering density at step n. The first term of the right-side of the equation is the observation equation whereas the second term represents the Chapman-Kolmogorov equation. Given the initialization at $n = 0$, the filter proceeds by solving the integral in the Chapman-Kolmogorov equation (prediction step) and compute the log posterior filtering density (update step). In our model, we solve Eq. (A.1) by means of a Gaussian approximation filter (Smith & Brown, 2003), which computes a gaussian approximation to the posterior density $f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta)$ and determines its posterior mode $z_{i,n|n}$ and variance $\lambda^2_{i,n|n}$ recursively.

More technically, let:

$$\log f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta) \propto \log f(y_{ijn}|z_{i,n},\theta)+ \tag{A.2}$$
$$+ \log \int_{\mathbb{R}} f(z_{i,n}|z_{i,n-1},\theta)f(z_{i,n-1}|\mathbf{y}_{ij,0:n-1},\theta)dz_{i,n-1}$$

be the filtering density at step n. Consider the following definitions:

$$y_{ijn}|z_{i,n} \sim mixVM(\mu_1,\mu_2,\kappa_1,\kappa_2,\pi_{ijn}) \tag{A.3}$$
$$z_{i,n}|z_{i,n-1} \sim \mathcal{N}(z_{i,n-1},\sigma_i^2) \tag{A.4}$$
$$z_{i,n-1}|\mathbf{y}_{ij,0:n-1} \sim \mathcal{N}(z_{i,n-1|n-1},\lambda^2_{i,n-1|n-1}) \tag{A.5}$$

where $z_{i,n-1|n-1}$ and $\lambda^2_{i,n-1|n-1}$ represent the mode and the variance of the gaussian approximation in the prediction step. Under definitions (A.4)-(A.5), integrating out for $z_{i,n-1}$ in Eq.(A.2) yields to:

$$z_{i,n}|\mathbf{y}_{ij,0:n-1},\theta \sim \mathcal{N}(z_{i,n-1|n-1},\lambda^2_{i,n-1|n-1}+\sigma_i^2) \tag{A.6}$$

For the sake of computational simplicity, we rewrite the measurement density in Eq. (2) as follows:

$$f(y_{ijn}|\pi_{ijn},\theta) = \left[\frac{\pi_{ijn}}{2\pi I_0(\kappa_1)}\exp\left(\cos(y_{ijn}-\mu_1)^{\kappa_1}\right)\right]^{u_{ijn}}\cdot$$
$$\cdot \left[\frac{1-\pi_{ijn}}{2\pi I_0(\kappa_2)}\exp\left(\cos(y_{ijn}-\mu_2)^{\kappa_2}\right)\right]^{1-u_{ijn}} \tag{A.7}$$

where u_{ijn} is a (deterministic) indicator variable taking the value 1 when y_{ijn} is in the area of the screen associated to *C1*, and 0 when y_{ijn} is in the area associated to *C2* (see Fig. 1) (Banerjee et al., 2005). Next, using these results together with Eqs. (A.3)-(A.5) we obtain:

$$\log f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta) \propto \log f(y_{ijn}|z_{i,n},\theta) + \log f(z_{i,n}|\mathbf{y}_{ij,0:n-1},\theta)$$

$$= u_{ijn}\big(\kappa_1\cos(y_{ijn}-\mu_1) + \log\pi_{ijn} - \log I_0(\kappa_1)\big) +$$

$$+ (1-u_{ijn})\big(\kappa_2\cos(y_{ijn}-\mu_2) + \log(1-\pi_{ijn}) - \log I_0(\kappa_2)\big) + \qquad \text{(A.8)}$$

$$+ \frac{1}{2}\left(\log(\lambda^2_{i,n-1|n-1} + \sigma_i^2) - \frac{(z_{i,n} - z_{i,n|n-1})^2}{\lambda^2_{i,n-1|n-1} + \sigma_i^2}\right)$$

Differentiate $\mathcal{F} \triangleq f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta)$ around $z_{i,n|n}$ gives:

$$\frac{\partial\log\mathcal{F}}{\partial z_{i,n|n}} = \sum_{j=1}^{J}\left(\frac{2z_{i,n|n-1} - 2z_{i,n|n}}{2\sigma_i^2 + \lambda^2_{i,n-1|n-1}} + \frac{\exp\big(\beta_j - z_{i,n|n}\big)u_{ijn}}{\exp\big(\beta_j - z_{i,n|n}\big) + 1} - \right.$$

$$\left. - \frac{\exp\big(\beta_j - z_{i,n|n}\big)(u_{ijn} - 1)}{\exp\big(2z_{i,n|n} - 2\beta_j\big) - 1}\right) \qquad \text{(A.9)}$$

$$\frac{\partial^2\log\mathcal{F}}{\partial z^2_{i,n|n}} = \sum_{j=1}^{J}\left(-\frac{1}{4\cosh\big(\frac{1}{2}(\beta_j - z_{i,n|n})\big)^2} - \frac{1}{(\sigma_i^2 + \lambda^2_{i,n-1|n-1})}\right) \qquad \text{(A.10)}$$

where $\cosh(x) \triangleq \big(1 + \exp(-2x)\big)/\big(2\exp(-x)\big)$ is the hyperbolic cosine function.
Finally, the posterior moment $z_{i,n|n}$ is obtained by solving Eq. (A.9) whereas $\lambda^2_{i,n|n}$ is computed by the negative inverse of Eq. (A.10) (e.g., see: Tanner, 1991). As Eq. (A.9) is non-linear, it can be solved numerically (e.g., using the Broyden's method). The complete filtering algorithm is summarized in Table A.1.

Algorithm 1	Gaussian Approximation filter algorithm	
n = 0 :	$z_{i,0} = 0$	INITIALIZATION
	$\lambda^2_{i,0} = 1$	
n > 0 :	$z_{i,n|n-1} = z_{i,n-1|n-1}$	PREDICTION
	$\lambda^2_{i,n|n-1} = \lambda^2_{i,n-1|n-1} + \sigma_i^2$	
	$z_{i,n|n} = \text{solve}\big(\partial\log f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta)/\partial z_{i,n|n}\big)$	UPDATE
	$\lambda^2_{i,n|n} = -\text{inv}\big(\partial^2\log f(z_{i,n}|\mathbf{y}_{ij,0:n},\theta)/\partial z^2_{i,n|n}\big)$	

Table A.1: Filtering algorithm on the interval $\{0, 1, \ldots, N\}$. The algorithm takes as input the parameters θ and the data $\{\mathbf{y}_{ij,0:N}, \mathbf{u}_{ij,0:N}\}$ whereas returns as output the filtered states $\mathbf{z}_{i,0:N}$. Note that the notation $n|n-1$ indicates the prediction of the current n-th state given the previous $n-1$ whereas $n|n$ denotes the update of the predicted state.

Finally, to ensure that the unobserved sequence $\mathbf{z}_{i,0:N}$ is an approximate realization from $f(\mathbf{z}_{i,0:N}|\mathbf{y}_{ij,0:N})$ we need to refine the filtering results conditional on the whole observed data $\mathbf{y}_{ij,0:N}$. This task is

achieved by means of a standard fixed-interval smoothing algorithm (Ansley & Kohn, 1982; Mendel, 1995), which uses the posterior filtering moments $z_{i,n|n}$ and $\lambda_{i,n|n}^2$ as input. The smoothing algorithm is described in Table A.2.

Algorithm 2	Fixed-interval smoothing algorithm	
$\mathbf{n = N}$:	$z_{i,N|N} \sim \mathcal{N}(z_{i,N|n}, \lambda_{i,N|n}^2)$	INITIALIZATION
	$\lambda_{i,N|N}^2 = 1$	
$\mathbf{n < N}$:	$z_{i,n|N} = z_{i,n|n} + \left(\dfrac{\lambda_{i,n|n}^2}{\lambda_{i,n+1|n}^2} \right)(z_{i,n+1|N} - z_{i,n+1|n})$	BACKWARD UPDATE
	$\lambda_{i,n|N}^2 = \lambda_{i,n|n}^2 + \left(\dfrac{\lambda_{i,n|n}^2}{\lambda_{i,n+1|n}^2} \right)^2 (\lambda_{i,n+1|N}^2 - \lambda_{i,n+1|n}^2)$	

Table A.2: Backward smoothing algorithm over the interval $\{N, N-1, \ldots, 0\}$. The algorithm takes as input the filtering solutions whereas returns as output the smoothed states $\mathbf{z}_{i,0:N}$ conditional on the whole set of data $\mathbf{y}_{ij,0:N}$. Note that the the notation $.|n$ refers to the filtering solutions whereas $.|N$ indicates the smoothing results.

Appendix B: Posteriors computation and estimation

In what follows, we describe the steps for computing the term $f(\boldsymbol{\Theta}|\mathbf{Y})$. First, we note that the array $\boldsymbol{\Theta}$ consists of two blocks of parameters associated to the observation equation and the stimuli equation of the model (3)-(4), namely J scalars $\{\beta_1, \ldots, \beta_J\}$ paired with the set of stimuli/trials and two parameters $\{\kappa_1, \kappa_2\}$ for the vonMises concentrations. To simplify the computations in the Metropolis-Hastings algorithm, the terms $\{\kappa_1, \kappa_2\}$ can be distinctively determined prior running the MH algorithm (Marin et al., 2005). Moreover, since we are not interested in the posterior distributions of these parameters, as long as they are not involved in the state-space dynamics, we compute them using the following Maximum-Likelihood estimators (Banerjee et al., 2005):

$$\hat{\kappa}_1 = I^{-1}\left(\frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{n=0}^N u_{ijn} \cos(y_{ijn} - \mu_1)}{\sum_{i=1}^I \sum_{j=1}^J \sum_{n=0}^N u_{ijn}} \right) \tag{A.11}$$

$$\hat{\kappa}_2 = I^{-1}\left(\frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{n=0}^N (1 - u_{ijn}) \cos(y_{ijn} - \mu_2)}{\sum_{i=1}^I \sum_{j=1}^J \sum_{n=0}^N (1 - u_{ijn})} \right) \tag{A.12}$$

where μ_1 and μ_2 are the location parameters fixed by the experimenter, u_{ijn} is defined as in Eq. (A.7), whereas I^{-1} is the inverse of the modified Bessel function which is evaluated numerically (Abramowitz et al., 1972). Given the above results and the constrains $\boldsymbol{\sigma}_{I \times 1} = \mathbf{1}_I$, the array of parameters simply reduces to $\boldsymbol{\theta}_{J \times 1}$. This simplifies the inner-working of the MH algorithm as it now works on a smaller and more compact parameter space. To proceed further, the decomposition (10) involves the following definition for the MH proposal density (Andrieu et al., 2010):

$$q(\{\boldsymbol{\theta}^{(t)}, \mathbf{Z}^{(t)}\} | \{\boldsymbol{\theta}^{(t-1)}, \mathbf{Z}^{(t-1)}\}) \; = \; q(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)}) f(\mathbf{Z}^{(t)} | \mathbf{Y}) \qquad \text{(A.13)}$$

where $f(\mathbf{Z}^{(t)} | \mathbf{Y})$ is evaluated through filtering/smoothing. This is appealing since the posterior density $f(\mathbf{Z}, \boldsymbol{\Theta} | \mathbf{Y})$ from which it might be difficult to sample from, reduces now to $f(\boldsymbol{\Theta} | \mathbf{Y})$ that is conveniently defined on a smaller parameters space (Andrieu et al., 2009). In our case, we can set:

$$q(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)}) \; = \; \mathcal{N}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\Sigma}^{(t)}) \qquad \text{(A.14)}$$

with $\boldsymbol{\Sigma}^{(t)}$ being a suitable $J \times J$ covariance matrix. Consequently, the MH acceptance ratio is as follows:

$$\alpha^{(t)} \; = \; \frac{f(\mathbf{Y} | \boldsymbol{\theta}^{(t)}) \; q(\boldsymbol{\theta}^{(t-1)} | \boldsymbol{\theta}^{(t)}) \; f(\boldsymbol{\theta}^{(t)})}{f(\mathbf{Y} | \boldsymbol{\theta}^{(t-1)}) \; q(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)}) \; f(\boldsymbol{\theta}^{(t-1)})} \qquad \text{(A.15)}$$

where $f(\mathbf{Y} | \boldsymbol{\theta})$ is the density for the marginal likelihood computation, $f(\boldsymbol{\theta})$ indicates the prior density over the parameters, whereas $q(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ is the MH proposal density.[5] Note that under (A.14), the terms $q(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ and $q(\boldsymbol{\theta}^{(t-1)} | \boldsymbol{\theta}^{(t)})$ in Eq. (A.15) vanish as they refer to the same probability value. This yields to:

$$\alpha^{(t)} \; = \; \frac{f(\mathbf{Y} | \boldsymbol{\theta}^{(t)}) \; f(\boldsymbol{\theta}^{(t)})}{f(\mathbf{Y} | \boldsymbol{\theta}^{(t-1)}) \; f(\boldsymbol{\theta}^{(t-1)})} \qquad \text{(A.16)}$$

where the MH ratio is now expressed as a function of the marginal likelihood and the priors. The term $f(\mathbf{Y} | \boldsymbol{\theta})$ can be easily computed as a byproduct of the filtering calculations (see Appendix C). Finally, the choice of a well-suited covariance matrix $\boldsymbol{\Sigma}^{(t)}$ for the proposal distribution is crucial in order to achieve chains' convergences. In our context, we used the Haario's adaptive solution where the proposal covariance is iteratively adapted during the chains using the current proposal covariance up to the adaptation step (Haario et al., 2001). The complete MH algorithm is summarized in Table A.3.

Appendix C: Marginal Likelihood computation

The marginal likelihood $f(\mathbf{Y} | \boldsymbol{\theta})$ is computed as follows. Let the observed-data log likelihood be:

$$\begin{aligned}
\log \mathcal{L}(\theta | \mathbf{Y}) &= \sum_{i=1}^{I} \sum_{j=1}^{J} \log f(\mathbf{y}_{ij,0:n} | \theta) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\sum_{n=0}^{N} \log f(y_{ijn} | \mathbf{y}_{ij,0:n-1}, \theta) \right) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\sum_{n=0}^{N} \log \int_{\mathbb{R}} f(y_{ijn} | z_{i,n}, \theta) f(z_{i,n} | \mathbf{y}_{ij,0:n-1}, \theta) \; dz_{i,n} \right)
\end{aligned} \qquad \text{(A.17)}$$

---

[5]Another possibility to define the proposal density for the Metropolis Hasting algorithm is that of using an efficient fully adaptive mixture proposal density for multi-modal and multi-dimensional target distributions as described by Luengo & Martino (2013).

Algorithm 3	Metropolis-Hasting algorithm	
$\mathbf{t} = \mathbf{0}:$	Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(0)}$, $\boldsymbol{\Sigma}^{(t)} = \boldsymbol{\Sigma}^{(0)}$	INITIALIZATION
	Run Algorithms 1-2 to get $\mathbf{Z}^{(0)} \sim f\big(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{(\mathbf{0})}\big)$	
	Compute $f(\mathbf{Y}|\boldsymbol{\theta}^{(\mathbf{0})})$	
$\mathbf{t} > \mathbf{0}:$	$\boldsymbol{\theta}^{\star} \sim \mathcal{N}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\Sigma}^{(t)})$	M-H LOOP
	Run Algorithms 1-2 to get $\mathbf{Z}^{\star} \sim f\big(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{\star}\big)$	
	Compute $f(\mathbf{Y}|\boldsymbol{\theta}^{\star})$	
	Compute $\alpha^{(t)}$ from Eq. (A.16)	
	Get $r \sim \mathcal{U}(0,1)$	ACCEPT/REJECT
	Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{\star}$, $\mathbf{Z}^{(t)} = \mathbf{Z}^{\star}$ if $\alpha^{(t)} \leq r$	
	Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$, $\mathbf{Z}^{(t)} = \mathbf{Z}^{(t-1)}$ if $\alpha^{(t)} > r$	
	Run $\boldsymbol{\Sigma}^{(t+1)} \leftarrow \text{adapt}\big(\boldsymbol{\Sigma}^{(t)}\big)$ see Haario et al. (2001)	ADAPTING PHASE

Table A.3: Marginal Metropolis-Hastings to estimate $\boldsymbol{\beta}_{J \times 1}$. Note that $\mathcal{U}(0,1)$ indicates the Uniform distribution over the interval $[0,1]$ whereas the adaptive phase can be performed at each iteration t of the chain or rather after a fixed interval $t + H$ (with $H > 1$).

According to the standard prediction error decomposition, this functional has been factorized as a function of both the measurement model density (2) and the one-step ahead predictive density (A.6) (De Jong, 1988). By substituting the above definitions with our model's densities, we get the following marginal likelihood:

$$
\log \mathcal{L}(\theta|\mathbf{Y}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{n=0}^{N} \log \Bigg(\int_{\mathbb{R}} \big(\pi_{ijn} \exp(\cos(y_{ijn} - \mu_1)^{\kappa_1})^{u_{ijn}} \cdot
$$
$$
\cdot \big((1 - \pi_{ijn}) \exp(\cos(y_{ijn} - \mu_2)^{\kappa_2})^{(1-u_{ijn})} \cdot
$$
$$
\cdot \exp\Big(-\frac{(z_{i,n} - z_{i,n|n-1})^2}{2(\sigma_i^2 + \lambda_{i,n-1|n-1}^2)} \Big) \, dz_{i,n} \Bigg)
\tag{A.18}
$$

where the von-Mises and Gaussian densities have been written by dropping the constant terms. Using the predictive moments $z_{i,n|n-1}$ and $\lambda_{i,n|n-1}^2$ from the filter algorithm (see Table A.1), a recursion on $\{0, 1, \ldots, N\}$ can be written for the likelihood computation, which consists of solving N integrals over the support of the r.v. $Z_{i,n}$. Since no analytical solutions are available for the functional above, numerical integration methods can be used to solve the integrals (e.g., see Shampine, 2008). Another possibility to compute the marginal likelihood is that of using a conditional Monte Carlo approximation by taking the advantages of the conditional equations of the models, in the spirit of Cabras et al. (2015).

Supplementary Materials

A state space approach to dynamic modeling of mouse-tracking data

Antonio Calcagnì[°][*], Luigi Lombardi[†], Marco D'Alessandro[†], Francesca Freuli[†]

[°]University of Padova, [†]University of Trento

Contents

[*]Corresponding author: antonio.calcagni@unipd.it

1 Evaluating the state-space model

1.1 Computation time

In this section we evaluate the time required by the algorithm included in Table A.3 of the manuscript to completely fit the state-space model. The algorithm has been implemented in Matlab R2018a and it is freely available online at `https://github.com/antcalcagni/SSM_mousetracking`. We tested two instances of our model, namely the simple model with a categorical variable and the complete model including categorical and continuous variables together with their interaction. The simulation study was obtained by varying the type of model (*categorical*, *interaction*), the number of statistical units $I = (12, 25, 50)$, the number of trials $J = (12, 28)$ and the number of levels of the categorical variable $K = (2, 4)$. The time step was kept fixed $N = 61$. The study involved $I \times J \times K = 12$ scenario for two type of models (see Table 1). For each scenario, data and model structure were kept fixed, one single chain was used without parallelization, warming-up and sampling periods were equal to 2000 samples (warm-up = 1000 samples), and the adaptive parameter of the MH-loop was equal to $H = 25$. The elapsed time was measured in seconds using the standard built-in Matlab function. Computations were performed on Intel Core i7-7500U 4-core 2.70GHz, Ubuntu 18.04 64-bit, 8GB Ram. Figure 1 shows the results for both types of models. As expected, the elapsed time increased exponentially as a function of the complexity of the scenario. In both models, the most simple scenario (scenario no. 1) required less then one minute to complete the warming-up and about one minute to complete the sampling period. Instead, the medium scenario (scenario no. 6) required about one minute to complete the warming-up and less then three minutes to complete the sampling phase. Finally, the most complex scenario (scenario no. 12) required less than three minutes to finalize the warming-up and about seven minutes to complete the sampling period.

scenario	parameters
1	I=12,J=12,K=2
2	I=12,J=12,K=4
3	I=12,J=28,K=2
4	I=12,J=28,K=4
5	I=25,J=12,K=2
6	I=25,J=12,K=4
7	I=25,J=28,K=2
8	I=25,J=28,K=4
9	I=50,J=12,K=2
10	I=50,J=12,K=4
11	I=50,J=28,K=2
12	I=50,J=28,K=4

Table 1: Computation time: Simulated scenario for both categorical and interaction models

1.2 Simulation study

In this section, we report the results of a simulation study to evaluate the ability of the model to (i) *recover* the true parameters $\boldsymbol{\beta}$ and (ii) *resemble* the observed data $\mathbf{Y}$. We designed the simulation under the Bayesian framework of statistical modeling. In particular, four factors were varied in a complete factorial design with 16 scenarios: type of model = {categorical, interaction}, $I = \{12, 25\}$, $J = \{12, 28\}$, $K = \{2, 4\}$. The combinations type of model $\times I \times J \times K$ were evaluated $M = 5000$

(A) Categorical model
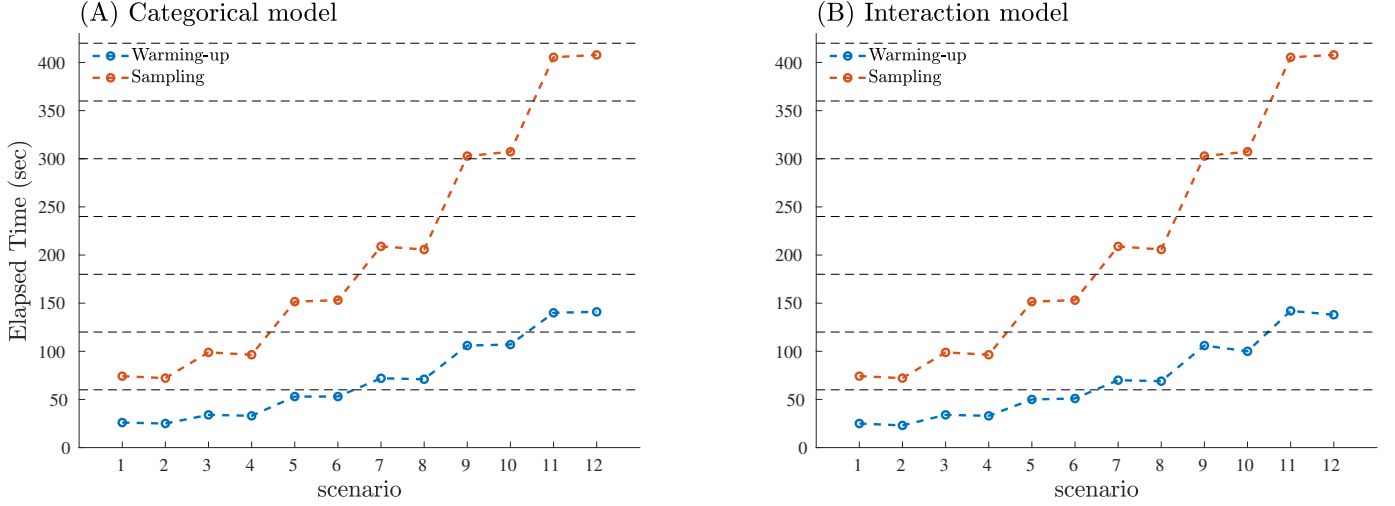
(B) Interaction model

Figure 1: Computation time: Elapsed time (sec.) with regards to warming-up (blue line) and sampling (red line) for the simplest categorical (A) and interaction (B) models. Note that dotted gray lines indicate minutes.

times by getting a total of 40000 final samples. The levels of I, J, and K were chosen to represent the number of individuals, trials, and variables commonly encountered in experimental psychology. The simulation was performed on a (remote) HPC machine based on 16 cpu Intel Xeon CPU E5-2630L v3 1.80GHz, 16x4GB Ram.

1.2.1 Simulation design

The simulation workflow proceeded as follows:

1. The hyperparameters $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$ of the prior $f(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ were kept fixed, $\mu_1 = 2.75$, $\mu_2 = 0.75$, $\kappa_1 = \kappa_2 = 100$, $X \sim \mathcal{U}_d(-3,3)$, $\mathbf{D} = \mathbf{I}_{K \times 1} \otimes \mathbf{1}_{\frac{J}{K} \times 1}$ (with $\otimes$ being the Kronecker product)

2. For each scenario $s = 1, \ldots, 16$, the parameters were sampled from the priors $\boldsymbol{\beta}_{1,\ldots,M} \sim f(\boldsymbol{\beta})$ and a series of M datasets $\mathbf{Y}_1^s, \ldots, \mathbf{Y}_M^s$ were generated according to the model equations (note that the generic dataset is of the form $\mathbf{Y}_{I \times J \times N}^s$)

3. For each of the datasets $\mathbf{Y}_1^s, \ldots, \mathbf{Y}_M^s$, the model was fitted by obtaining the array of latent states and parameter posteriors:

$$\mathcal{R} = \{(\mathbf{Z}_1^s, \ldots, \mathbf{Z}_M^s), (f(\boldsymbol{\beta}|\mathbf{Y})_1^s, \ldots, f(\boldsymbol{\beta}|\mathbf{Y})_M^s)\}$$

where the generic array $\mathbf{Z}^s$ is of the form $\mathbf{Z}_{I \times N}^s$. The model was run using 10000 (samples) x 8 (chains), with warming-up of 2500 samples. Matlab Parallel computing was used to run the chains.

1.2.2 Outcome measures

The *recovery problem* was evaluated for $s = 1, \ldots, 16$ by means of the overlapping measure between symmetric distributions:

$$\Lambda\Big(f(\boldsymbol{\beta})^s, f(\boldsymbol{\beta}|\mathbf{Y})^s_{1,\ldots,M}\Big) = \frac{1}{M} \sum_{m=1}^{M} \int_{\mathbb{R}^n} \min\Big(f(\boldsymbol{\beta})^s, f(\boldsymbol{\beta}|\mathbf{Y})^s_m\Big)$$

which measures the amount of agreement between two distributions, with $\Lambda = 0$ indicating no overlapping. In our context, as $f(\boldsymbol{\beta})$ and $f(\boldsymbol{\beta}|\mathbf{Y})$ are symmetric with a unique mode, we interpret Λ as density similarity where $\Lambda = 1$ means that the two distributions are the same[1].

The *resembling problem* was instead assessed by means of the $\mathbf{Y}|\tilde{\mathbf{Y}}$-discrepancy measure:

$$\overline{\mathrm{AoR}}^s = \frac{1}{M} \sum_{m=1}^{M} \mathrm{AoR}\Big(\mathbf{Y}^s_m, \tilde{\mathbf{Y}}^s_m\Big)$$

where:

$$\mathrm{AoR}\Big(\mathbf{Y}, \tilde{\mathbf{Y}}\Big) = 100 \cdot \frac{\min\big(||\mathbf{y}^\dagger||, ||\mathrm{vec}(\mathbf{Y})||\big)}{\max\big(||\mathbf{y}^\dagger||, ||\mathrm{vec}(\mathbf{Y})||\big)} \quad \text{with} \quad \mathbf{y}^\dagger = \frac{\mathrm{vec}(\tilde{\mathbf{Y}})^T \mathrm{vec}(\mathbf{Y})}{||\mathrm{vec}(\mathbf{Y})||} \cdot \frac{\mathrm{vec}(\mathbf{Y})}{||\mathrm{vec}(\mathbf{Y})||}$$

which measures the *amount of reconstruction*[2] of the observed array $\mathbf{Y}$ via the reproduced array $\tilde{\mathbf{Y}}$. The measure AoR takes values in [0,100] with 0% indicating that no reconstruction occurred at all.

1.2.3 Results

The results of the simulation study are described in Tables 2-3. Overall, both the models got acceptable fit with regard to recover the true model structure (Λ index) and the adequacy in reproducing the input data (AoR index). As expected, the goodness-of-fit increased as a function of the number of trials J and the number of individuals I.

		AoR	Λ_{γ_1}	Λ_{γ_2}	Λ_{γ_3}	Λ_{γ_4}
	$J = 12, K = 2$	0.688	0.816	0.790		
	$J = 12, K = 4$	0.746	0.825	0.862	0.853	0.794
$I = 12$	$J = 28, K = 2$	0.732	0.828	0.848		
	$J = 28, K = 4$	0.799	0.874	0.807	0.830	0.831
	$J = 12, K = 2$	0.780	0.848	0.796		
	$J = 12, K = 4$	0.808	0.749	0.841	0.805	0.895
$I = 25$	$J = 28, K = 2$	0.753	0.809	0.844		
	$J = 28, K = 4$	0.811	0.828	0.866	0.833	0.857

Table 2: Simulation study (Categorical model): AoR and Λ indices for each level of the simulation design.

---

[1]Schmid, F., & Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping—theory and empirical application. *Computational statistics & data analysis*, 50(6), 1583-1596.

[2]Sadtler, P. T., Quick, K. M., et al. (2014). Neural constraints on learning. *Nature*, 512(7515), 423-426.

		AoR	Λ_{γ_1}	Λ_{γ_2}	Λ_{γ_3}	Λ_{γ_4}	Λ_η	Λ_{δ_1}	Λ_{δ_2}	Λ_{δ_3}
	$J = 12, K = 2$	0.695	0.809	0.843			0.637	0.841		
	$J = 12, K = 4$	0.720	0.909	0.880	0.943	0.754	0.685	0.794	0.872	0.827
$I = 12$	$J = 28, K = 2$	0.807	0.813	0.831			0.684	0.835		
	$J = 28, K = 4$	0.710	0.955	0.946	0.863	0.902	0.674	0.933	0.888	0.900
	$J = 12, K = 2$	0.751	0.816	0.843			0.737	0.789		
	$J = 12, K = 4$	0.861	0.919	0.893	0.885	0.910	0.675	0.838	0.897	0.835
$I = 25$	$J = 28, K = 2$	0.794	0.862	0.842			0.701	0.762		
	$J = 28, K = 4$	0.873	0.875	0.848	0.981	0.806	0.726	0.887	0.870	0.869

Table 3: Simulation study (Interaction model): AoR and Λ indices for each level of the simulation design.

2 Case study

2.1 Convergences of the MCMC chains

In this section we illustrate the convergence diagnostics for the three case studies described in Section 4 of the manuscript.
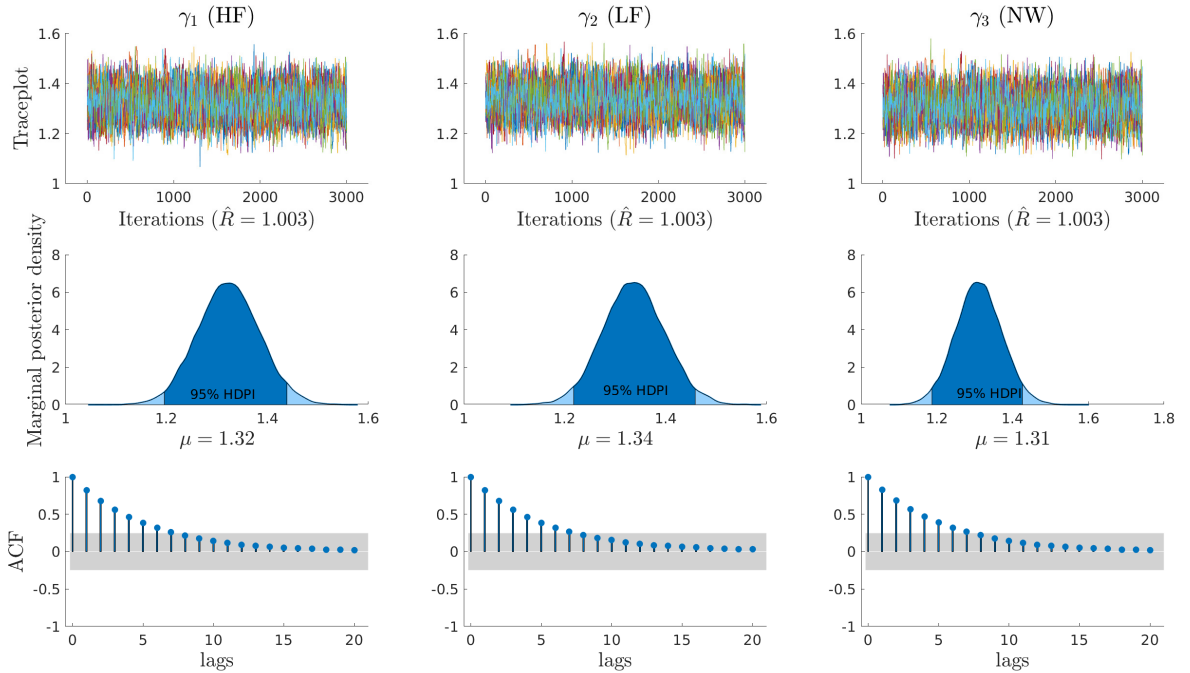
Figure 2: Case study (Model 2): MCMC diagnostics. Left panel: Traceplot of the chains in different colors and Gelman-Rubin $\hat{R}$ index. Middle panel: Marginal posterior density with posterior means μ and 95% HDPI. Right panel: Autocorrelation function (ACF) for the collapsed chains.
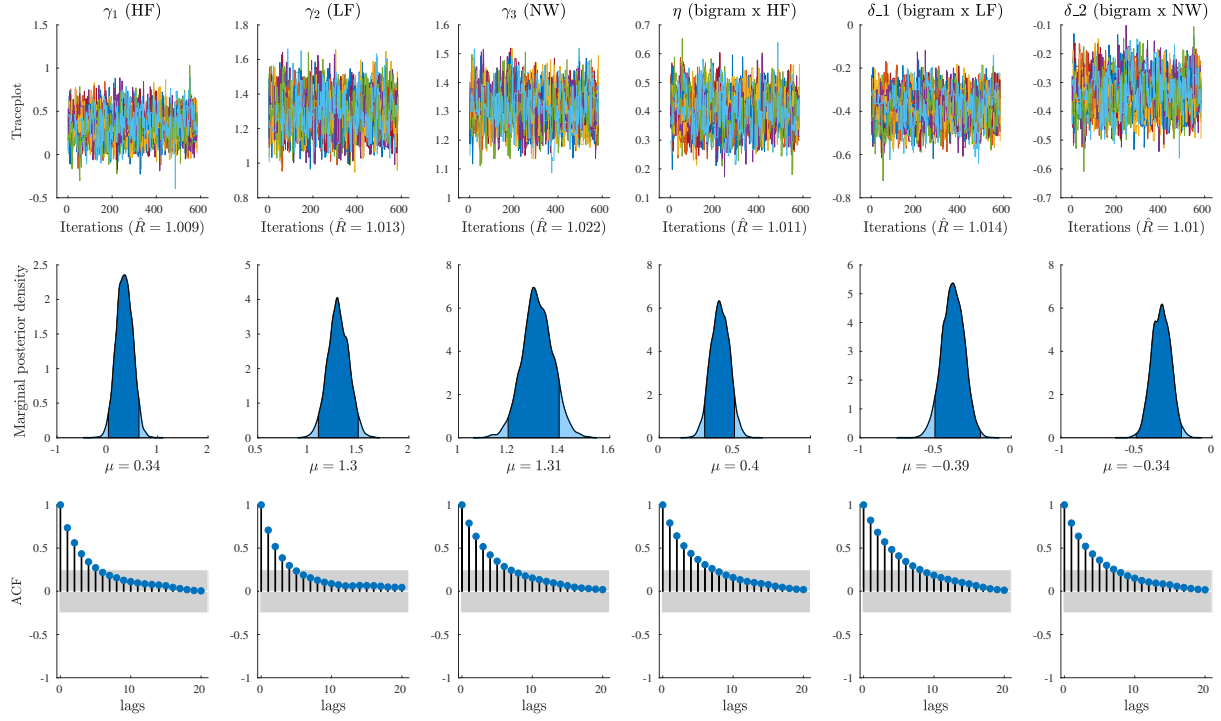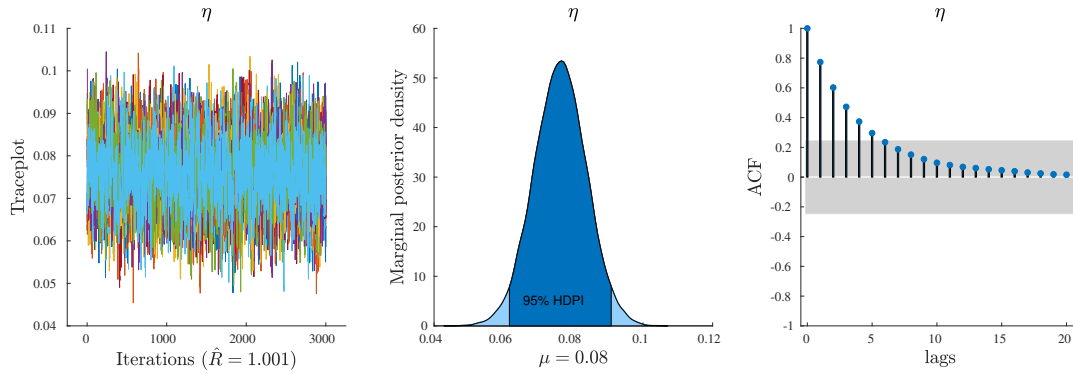
5

Figure 3: Case study (Model 1): MCMC diagnostics. First row: Traceplot of the chains in different colors and Gelman-Rubin $\hat{R}$ index. Second row: Marginal posterior densities with posterior means μ and 95% HDPI. Third row: Autocorrelation function (ACFs) for the collapsed chains.

Figure 4: Case study (Model 3): MCMC diagnostics. First row: Traceplot of the chains in different colors and Gelman-Rubin $\hat{R}$ index. Second row: Marginal posterior densities with posterior means μ and 95% HDPI. Third row: Autocorrelation function (ACFs) for the collapsed chains.

2.2 Fit of the model

The adequacy of the state-space model to reproduce the observed data was computed by means of a simulation-based approach. In particular, given the posteriors of the parameters $\hat{\boldsymbol{\beta}}$ and filtered/smoothed latent states $\hat{\mathbf{Z}}$, M new (simulated) datasets $(\mathbf{Y}_1^*, \ldots, \mathbf{Y}_M^*)$ were generated according to the estimated model structure. For each new dataset, the AoR discrepancy measure was computed:

$$\text{AoR} = 100 \cdot \frac{\min\left(||\mathbf{y}^\dagger||, ||\text{vec}(\mathbf{Y})||\right)}{\max\left(||\mathbf{y}^\dagger||, ||\text{vec}(\mathbf{Y})||\right)} \quad \text{with} \quad \mathbf{y}^\dagger = \frac{\text{vec}(\mathbf{Y}_m^*)^T \text{vec}(\mathbf{Y})}{||\text{vec}(\mathbf{Y})||} \cdot \frac{\text{vec}(\mathbf{Y})}{||\text{vec}(\mathbf{Y})||} \qquad m = 1, \ldots, M$$

In addition, to further evaluate the amount of reconstruction, we computed two kind of AoR measures:

- **Overall AoR:** percentage of data reconstruction based on the overall $I \times J \times N$ observed matrix $\mathbf{Y}$.

- **By-subject AoR:** percentage of data reconstruction based on the $J \times N$ observed matrix $\mathbf{Y}_i$ for each subject $i = 1, \ldots, I$. The index allows for evaluating the adequacy of the model to reconstruct the individual-based set of trajectories.

The simulation was performed on a (remote) HPC machine based on 16 cpu Intel Xeon CPU E5-2630L v3 1.80GHz, 16x4GB Ram. Figures 5-7 show the results of AoR indices for the three models. Figure 8 shows the observed mean trajectories for each individual $\mathbf{Y}_i$, $i = 1, \ldots, I$ against the predicted trajectories by the three models.
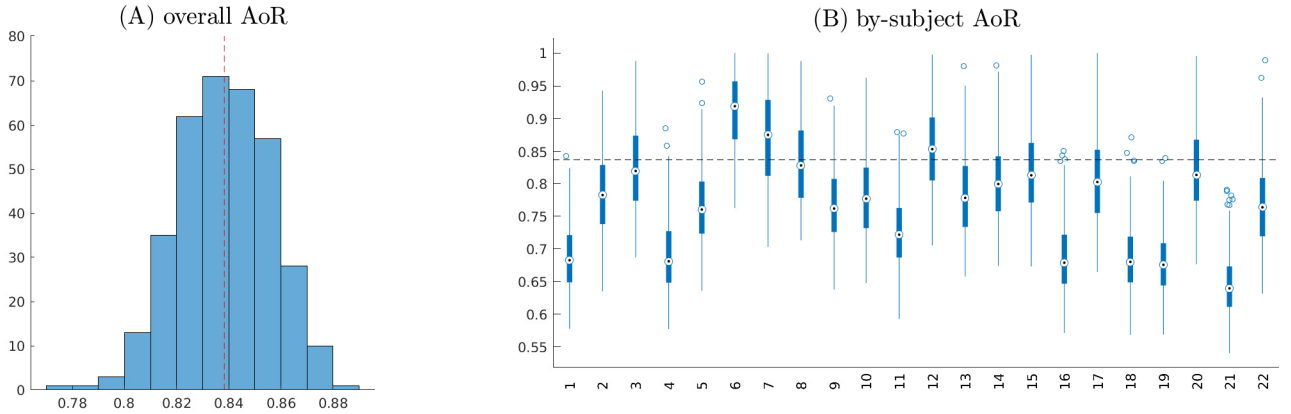

Figure 5: Goodness of fit (Model 1): (A) Overall amount of reconstruction and (B) By-subject amount of reconstruction. Note that dotted lines represent the mean of the overall AoR index.
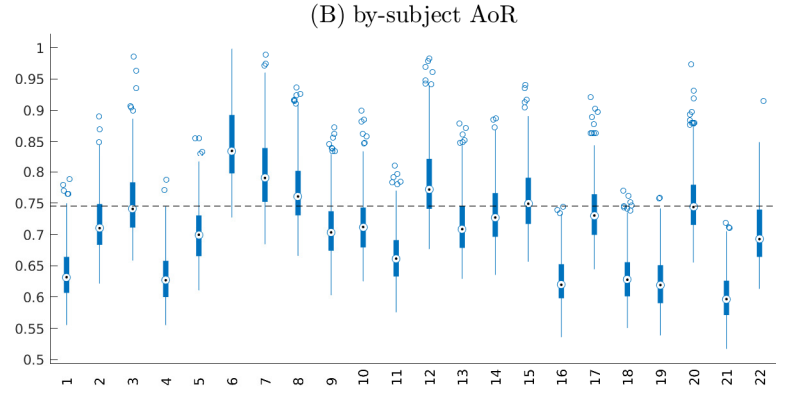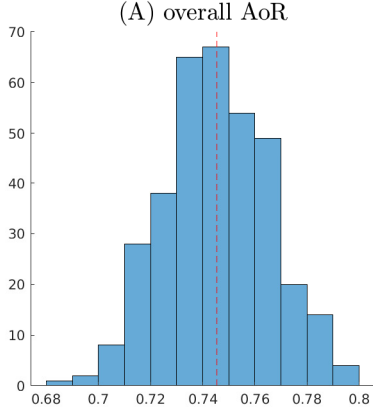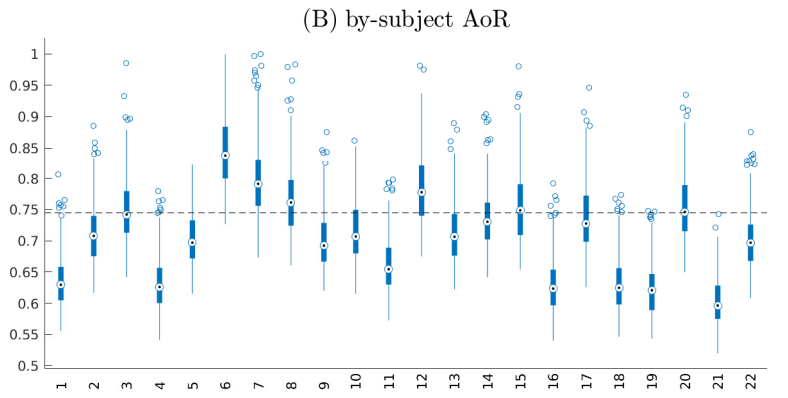
(A) overall AoR

(B) by-subject AoR

Figure 6: Goodness of fit (Model 2): (A) Overall amount of reconstruction and (B) By-subject amount of reconstruction. Note that dotted lines represent the mean of the overall AoR index.
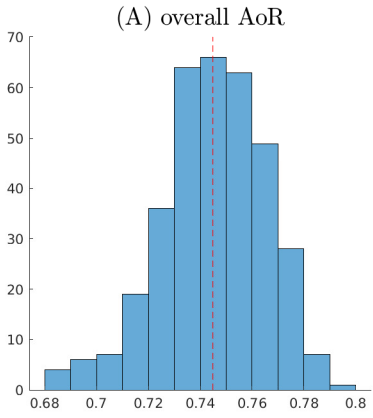

(A) overall AoR

(B) by-subject AoR

Figure 7: Goodness of fit (Model 3): (A) Overall amount of reconstruction and (B) By-subject amount of reconstruction. Note that dotted lines represent the mean of the overall AoR index.
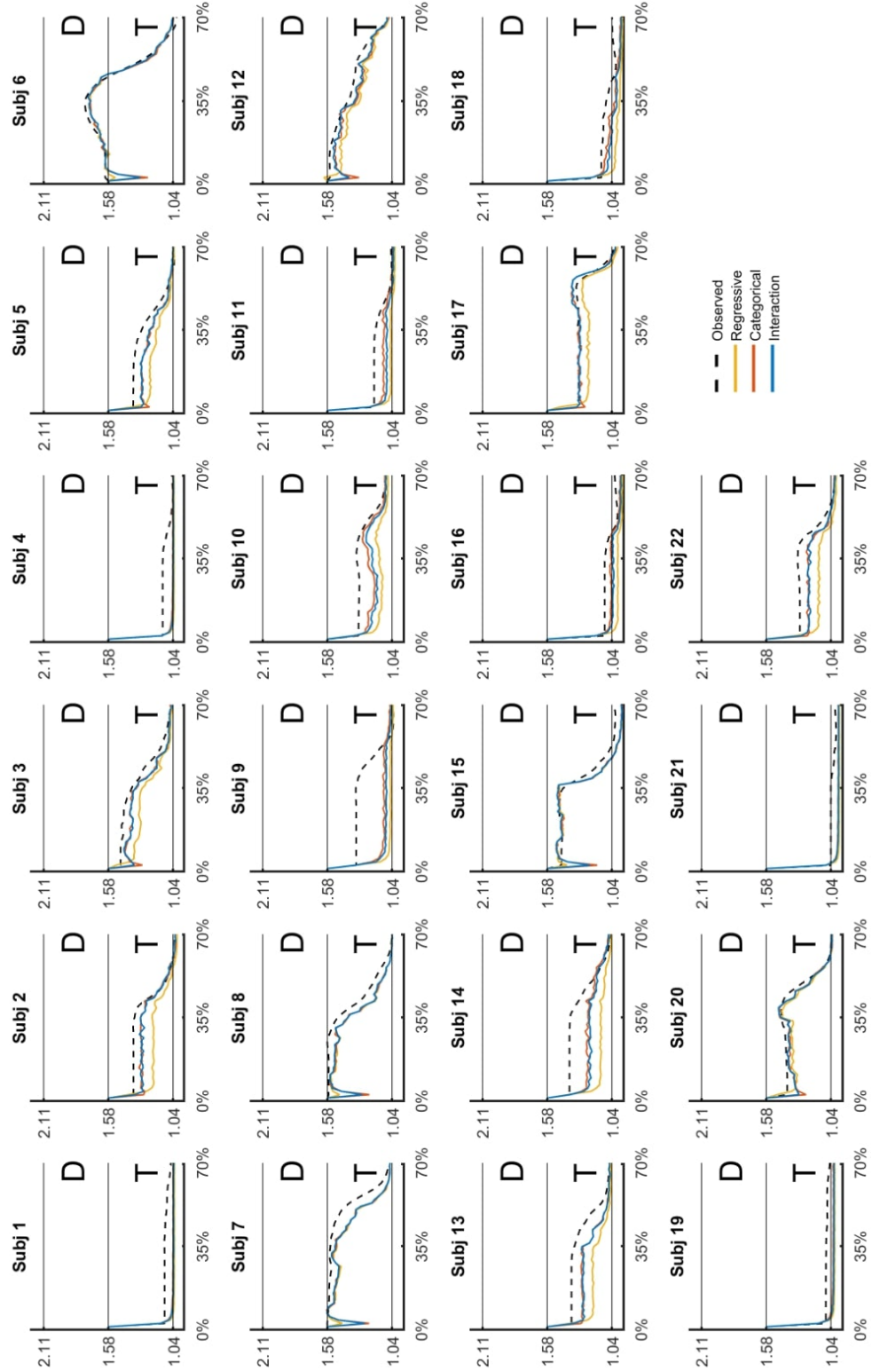
8

Figure 8: Goodness of fit: Observed and predicted individual (mean) trajectories for each of three models. Note that dotted lines represent observed trajectories whereas the upper and lower halves indicate distractor (D) and target (T) hemispaces, respectively.

9