2020, Computational Brain & Behavior, 3 (4), 430-457
https://link.springer.com/article/10.1007/s42113-020-00072-0

Extensions of Multivariate Dynamical Systems to Simultaneously Explain Neural and Behavioral Data

Qingfang Liu[a], Alexander A. Petrov[a], Zhong-Lin Lu[b,c], Brandon M. Turner[a,*]

[a] *Department of Psychology, The Ohio State University, Columbus, USA*
[b] *Division of Arts and Sciences, NYU Shanghai, Shanghai, China*
[c] *Center for Neural Science and Department of Psychology, New York University, New York, USA*

Abstract

To examine how the brain produces behavior, new statistical methods have linked neurophysiological measures directly to mechanisms of cognitive models, modeling both modalities simultaneously. However, current simultaneous modeling efforts are largely based on either correlational methods or on functions that map one stream of data to the other. Such frameworks are limited in their ability to infer causality between brain activity and behavior, typically ignore important temporal dynamics of neural measures, or ignore large and small scale functional networks necessary for completing cognitive tasks. In this article, we investigate one causal framework for modeling brain dynamics as a potential alternative for explaining how behavior can be viewed as an emergent property of brain dynamics. Our proposed framework can be considered an extension of Multivariate Dynamical Systems (MDS; Ryali et al., 2011), as it is constructed in a way such that the temporal dynamics and brain functional connectivities are explicitly contained in the model structures. To test the potential usefulness of the MDS framework, we formulate a concrete model within it, demonstrate that it generates reasonable predictions about both behavioral and fMRI data, and conduct a parameter recovery study. Specifically, we develop a generative model of perceptual decision making in a visual motion direction discrimination task.

*Corresponding Author
Email address: turner.826@gmail.com (Brandon M. Turner)
This research was supported by NSF-SMA 1533500

Two simulation studies under different experimental protocols illustrate that the MDS model can capture key characteristics of both behavioral and neural measures that typically occur in experimental data. We also examine whether or not such a complex system can be inferred from experimental data by evaluating whether current algorithms for fitting models to data can recover sensible parameter estimates. Our parameter recovery study suggests that the MDS parameters can be recovered using likelihood-free estimation techniques. Together, these results suggest that our MDS-based framework shows great promise for developing fully integrative models of brain-behavior relationships.

Keywords: Joint modeling, Dynamical systems, Bayesian inference, Perceptual decision making

---

1. Introduction

The rapid development of brain measurement techniques such as functional magnetic resonance imaging (fMRI) have contributed substantial insights into the neural correlates of human information processing and cognitive operations in cognitive neuroscience. Traditional cognitive neuroscience has investigated relations between brain and behavior in two directions. The first direction is on interpreting and understanding the unique contribution of individual brain areas, known as localization. The central premise is that different brain areas are different because they perform different operations. For example, certain brain regions (e.g. V5 or middle temporal (MT)) are thought to play major roles in processing visual motion (Maunsell and Van Essen, 1983; Vanduffel et al., 2001), in that the neurons in these regions significantly predict decisions about motion direction (Gold and Shadlen, 2007). The second direction is to identify brain networks that jointly describe cognitive operations, where the premise is that the completion of any cognitive function requires the collaboration of a series of functionally segregated brain functions. For example, in the case of visual motion processing, the completion of the function also relies on some basic cortical or subcortical functions such as the basal ganglia to either inhibit the motion impulse or execute a motor command (Hikosaka et al., 2000a; Lo and Wang, 2006). Both directions contribute to our understanding of how individual brain regions work together within a functional network to produce behavior, and what the functional roles of those individual brain regions are within the context of a task. However, most analyses in cognitive neuroscience consider the questions of "what is the functional role of brain region X?" and "what is the brain network that gives rise to cognitive operation Y?" as two separate issues, often requiring completely different statistical techniques. Segregating these two objectives can potentially obfuscate the functional interpretation of brain region X, specifically what its contribution to cognitive operation Y actually is.

To better understand and interpret brain function, a new wave of researchers have abstracted away the cognitive operations necessary for performing cognitive tasks, and examined how these abstractions are related to brain activity (see Forstmann and Wagenmakers, 2015; Turner et al., 2017b, 2019a,b; de Hollander et al., 2016, for reviews). These efforts are based on a set of linking propositions (Teller, 1984; Schall, 2003) relating psychological variables to physiological ones, where various approaches can be uniquely

separated on the basis of how researchers impose said link (de Hollander et al., 2016; Turner et al., 2017b). Although a detailed review is beyond the purpose of this article, Fig. 1 shows a few particularly relevant diagrams that illustrate different linking concepts within an overarching "joint modeling" framework (Turner et al., 2013b, 2015b; Palestro et al., 2018a; Turner et al., 2019a). The directed approach (left) attempts to simply transform the neural data N into a parameter θ within a cognitive model, and the transformation may have parameters δ. The success of this linking procedure is the degree to which a suitable transformation of the neural data provides good predictions for, or fits to, behavioral data B. The covariance approach (middle) attempts to impose a flexible map from neural data to model parameters by considering all possible pairwise correlations between sets of candidate brain regions and mechanisms in the model. It assumes an overarching distribution that enforces an explicit connection between parameters θ, δ, and Ω, where δ and θ have a direct constraint on neural N and behavioral B data, respectively. Although new efforts have increased the scalability of this approach (Turner et al., 2017a), there are clear limitations with considering all possible pairwise correlations.

The two approaches – directed and covariance – each attempt to address both the functional role of brain regions and the overarching functional network among brain regions. The directed approach instantiates an explicit link between neural data from one brain region and a model parameter. Once fit to data, one can then assess the degree to which a significant mapping relation exists by, for example, examining the posterior distribution of the slope parameter within a regression model linking N to θ. Although an informed multivariate regression model is a possible solution, directed models are not typically made cognizant of the many interactions that may exist between different brain regions, and hence are typically not suitable for uncovering brain networks. On the other hand, the covariance approach was intended to extract brain networks by considering the set of brain regions that (1) are correlated with one another, and (2) are jointly correlated with a cognitive mechanism. Despite the promise of network extraction, covariance models are still limited in the sense that they are typically correlational in nature. The linking function most commonly prescribed is a multivariate normal distribution (e.g., Turner et al., 2015b, 2016), such that the connections among modalities are defined by a covariance matrix. Directed models impose a more causal, confirmatory structure (Cassey et al., 2016; van Ravenzwaaij et al., 2017) but are also tied to specific details of cognitive models that may

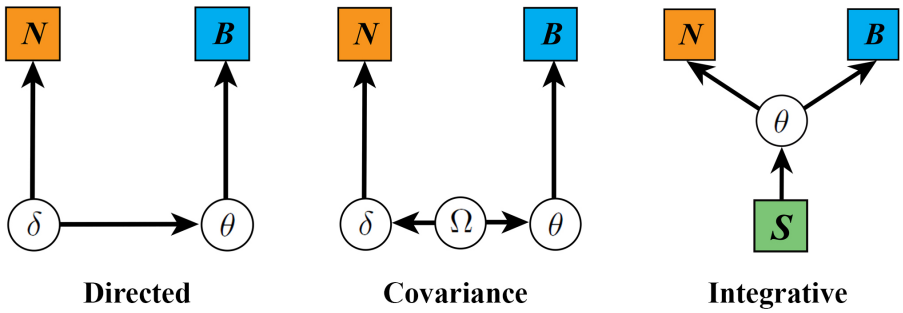

Figure 1: **An illustration of three approaches for linking neural and behavioral data simultaneously.** N represents the neural data, B represents the behavioral data, and S represents experimental stimuli. θ, δ, and Ω represent model parameters. Solid lines with arrows indicate ancestry statistical dependence among the nodes in the graph.

limit their flexibility.

In this article, we explore a different approach that we refer to as the "integrative" approach shown in the right panel of Fig. 1. The integrative approach develops a single cognitive model capable of predicting both neural and behavioral measures from experimental stimuli S. Here, a single set of parameters θ transforms the experimental stimuli through a model specification to generate predictions about neural N and behavioral B data jointly. Integrative models have been previously developed and productively used. For example, Anderson and colleagues (Anderson, 2007; Anderson et al., 2008; Borst and Anderson, 2013; Borst et al., 2010a,b) have shown that by using the ACT-R architecture to specify the model structure (i.e., θ in Fig. 1), fMRI data can be predicted by convolving modular activation within ACT-R with the canonical hemodynamic response function. Because ACT-R was already designed to explain behavioral data, the internal dynamics of ACT-R give a natural mechanism for also producing predictions for neural data. The integrative approach is also related to the work of Cassey et al. (2016), Kragel et al. (2015), Polyn et al. (2005), and Purcell et al. (2010), where neural data are modeled and directly drive or replace components of a cognitive model.

Although conceptually simple in Fig. 1, the success of an integrative model is determined by how the model structure in θ is specified. Our goal in this article is to create a framework for designing integrative models, by (1) identifying key brain regions that jointly contribute to the cognitive processes in the task, (2) defining structure among those brain regions that respects the temporal and spatial properties of brain regions having a physical existence within space and time, and (3) specifying how activity in a subset of brain regions promotes a specific behavioral response. Our framework considers the distributed interactions among brain regions by conceptualizing them as being temporally and spatially dependent, yet functionally integrated (Friston, 2009). To provide constraint on integrative models, we articulate our framework by requiring full specification of the time series for each region of interest (ROI). The time series of each ROI will be a dependent function of all brain regions in the set, which will allow us to investigate both localization behavior and functional connectivity among ROIs, potentially providing an integrated solution to understanding the functional role of ROIs within a network.

By virtue of their specificity, integrative models, with the form shown in Fig. 1, are difficult to develop and fit to data. Not only must integrative models consider how brain regions interact with one another, they must also consider how those regions ultimately give rise to a prediction about behavior. Often, researchers can rely on previous localization work to define how brain regions contribute to the cognitive process, but this is no small task, especially considering the emergence of brain networks with common functional structure discussed above. Also, there are methodological difficulties in fitting integrative models to data because they have a larger number of parameters and they often are mathematically intractable due to their inherently stochastic and time-dependent nature.

We propose a new integrative framework for mapping functional brain activity to decision making processes, based on Multivariate Dynamical Systems (MDS; Ryali et al., 2011). Our framework is designed to simultaneously generate behavioral data and neural measures for cognitive tasks. In constructing this framework, we have three criteria in mind. First, our framework should construct fully generative models for neuro-cognitive processes. Generative models predict the pattern of neural and behavioral data a priori based on assumptions of underlying cognitive processes and stimulus properties. Second, our framework should explain neural measures from a functionally integrated brain network, such that the coordination contributes

to the eventual cognitive process. Third, we wish to specify the generative process for neural data in an abstract, measure-independent space such that integrative models are invariant with respect to the type of neural measures collected experimentally (e.g., fMRI, EEG). Imposing these constraints here will facilitate future work enabling data fusion, where a single cognitive model can be used to explain behavior, EEG, and fMRI (e.g., Turner et al., 2016).

In this article, we use our framework to construct a specific cognitive model for the perceptual decision making task. We present two simulation results showing that the extended MDS models can generate plausible patterns of both behavioral and neural data. We then investigate whether or not such a framework can be realized from neural and behavioral data from a cognitive task. To investigate this, we apply approximate Bayesian methods to estimate model parameters of the model by fitting it to simulated data. Finally, contributions and limitations of the extended MDS are discussed.

2. Multivariate Dynamical Systems

The proposed MDS framework is closely related to but also distinct from certain other frameworks. On the one hand, MDS can be viewed as a multivariate version of the linear dynamical systems. For example, bilinear dynamical systems model a single neuron activation (Penny et al., 2005), and switching linear dynamic systems are proposed to improve the overall quality and sufficiency of model parameter estimation (Smith et al., 2010). On the other hand, MDS has many commonalities with dynamic causal modeling (DCM; Friston et al., 2003, Friston et al., 2017, Marreiros et al., 2008, Stephan et al., 2010) in that they both contain a "state equation" to model the latent neuronal activations, and an "observation equation" to map the latent neuronal activation to the observed neural signals, such as fMRI blood oxygen level dependent (BOLD) signals. However, there are many differences between MDS and DCM. First, conventional DCM treats the brain as a deterministic dynamic system subject to inputs (Friston et al., 2003) although a stochastic DCM was developed later (Daunizeau et al., 2009), whereas MDS explicitly includes a stochastic term. Second, DCM and MDS use different observation equations to map the latent neuronal activation to the BOLD signal. In particular, DCM adopts a nonlinear "Balloon" model (Buxton et al., 1998; Friston et al., 2000; Mandeville et al., 1999; Stephan et al., 2007) to describe how latent neuronal activations are transformed into hemodynamic time-series, while MDS formulates the relationship as a linear convolution of

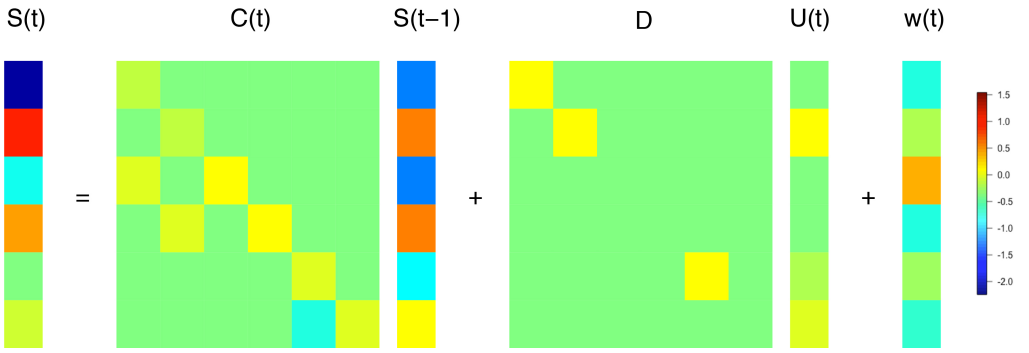

Figure 2: **An illustration of the MDS state equation for a model with 6 ROIs.** $S(t)$ and $S(t-1)$ are two column vectors denoting the neuronal activations at time point t and $t-1$, respectively. $C(t)$ is a endogenous brain connectivity matrix at time t. D is a diagonal matrix with direct exogenous effects indicated by diagonal entries. $U(t)$ is the strength of input. $\omega(t)$ denotes the noise vector.

latent neuronal states with a kernel expansion using basis functions (Ryali et al., 2011). Another DCM study related to the goal of the current article is the behavioral DCM (Daunizeau et al., 2014; Rigoux and Daunizeau, 2015). The central idea of the behavioral DCM is that the hidden neuronal states can be transformed by a probabilistic sigmoid mapping to produce a binary behavioral choice. Our extended MDS model can produce both behavioral choice and response time, under a mechanistic model.

In general, MDS is a state-space model in that it models observed data by assuming a time series of unobserved data. MDS first captures unobserved states by specifying a state equation, and then maps the unobserved states to observed data by specifying an observation equation. Here we consider neural measures as observed data and consider neuronal activations in the brain ROIs as the latent states. An important point that will be exemplified in future studies is that while neural measures are directly dependent on the measurement tools (e.g. fMRI BOLD signals, EEG signals), the latent neuronal activations in this framework are invariant to the measurement tools. Hence, once the latent activations are specified in a given system, any number of neural measures may be used to infer the parameters of the model from data.

2.1. State Equation

For M brain ROIs, we denote ROI i as R_i $(i = 1, \ldots, M)$. $S(t)$ represents neuronal activations at time t in each of M ROIs and it is a column vector

of length M. The MDS state equation

$$S(t) = C(t)S(t-1) + DU(t) + \omega(t), \text{for } t = 1, \ldots, T, \quad (1)$$

involves a sum of three terms that are illustrated in Fig. 2 for a model with 6 ROIs. First, $C(t)$ is an $M \times M$ matrix showing the strengths of endogenous brain connectivity at time point t. The diagonal elements in $C(t)$ indicate the self-connection within each ROI, and non-diagonal elements indicate interconnection paths between ROIs. For example, $C[3, 1]$ denotes the connectivity strength from R_1 to R_3, and this connectivity could differ from $C[1, 3]$, the connectivity strength from R_3 to R_1. Notice that $C(t)$ is often assumed to be time-invariant in dynamical systems, but here we allow this matrix to vary across time to accommodate the specifications of our cognitive model.

Second, the term $DU(t)$ in Eq. 1 indicates the direct exogenous effect on $S(t)$. The vector $U(t)$ has M components, each of which indicates the strength of external inputs to the corresponding ROI at time t. The strength values are mainly affected by the experimental stimuli property. $U(t)$ can be constant across time T but can also vary to represent temporal fluctuations of the perceived strength values. D is an $M \times M$ diagonal matrix and $D(i, i)$ weights the external inputs. By specifying a diagonal matrix, each external stimulus is constrained to affect exactly one ROI.

Third, the noise term $\omega(t)$ is a vector of length M sampled from a multivariate normal distribution with $\omega(t) \sim N_M(0, Q(t))$. This most general form of the variance-covariance matrix $Q(t)$ indicates that noise can vary across time and may be correlated across different ROIs. If one assumes independent and identically distributed noise across both time and ROIs, then $Q(t)$ can be simplified to $\sigma^2 I_M$, where I_M is an identity matrix of size M (Ryali et al., 2011). $Q(t)$ is essentially useful as a way to manipulate the signal-to-noise ratio (Ryali et al., 2011), and thus in our framework, $Q(t)$ systematically affects choice accuracy and response times. For the simulation in this article, $Q(t)$ can vary across time in order to accommodate the specifications of our cognitive model.

2.2. Observation Equation

We choose fMRI BOLD signal as the neural measure for the purposes of this article. In MDS, the BOLD signal in each ROI is modeled as a linear convolution of the hemodynamic response function (HRF) and latent neuronal activations in each ROI with appropriate observation noise (Ryali

et al., 2011). The latent neuronal activation in R_m at time t comes from the m-th element of $S(t)$ and is denoted as $S_m(t)$. The observed BOLD signal at time t in R_m is denoted as $Y_m(t)$. If we use $h_m(\tau)$ to denote the impulse response, or the HRF for R_m, the observation equation can be expressed as

$$\begin{aligned} Y_m(t) &= S_m(t) \otimes h_m(\tau) + e_m(t) \\ &= \int_{-\infty}^{\infty} S_m(t-\tau) h_m(\tau) d\tau + e_m(t). \end{aligned} \tag{2}$$

where "$\otimes$" denotes linear convolution, and $e_m(t)$ is the observation noise. The subscript m in each component allows regional variability. Here, we assume that $h_m(\tau)$ takes the canonical form of the double gamma model implemented in SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/):

$$h_m(\tau) = A_m \left[\frac{\tau^{\alpha_1 - 1} \beta_1^{\alpha_1} e^{-\beta_1 \tau}}{\Gamma(\alpha_1)} - c \frac{\tau^{\alpha_2 - 1} \beta_2^{\alpha_2} e^{-\beta_2 \tau}}{\Gamma(\alpha_2)} \right], \text{for m} = 1, \ldots, M, \tag{3}$$

where τ references time and $\Gamma(x) = (x-1)!$ indicates the Gamma function, which acts as a normalization term. By convention, we set $\alpha_1 = 6$, $\alpha_2 = 16$, $\beta_1 = \beta_2 = 1$ and $c = 1/6$ to represent the shape of HRF. The unknown parameter in the HRF is the amplitude A_m, dependent on ROI R_m. The other unknown aspect of Equation 3 is the length of the HRF (denoted as L, in seconds). We choose to produce neuronal activations $S_m(t)$ on the millisecond level, and so we set $\tau = \{.001, .002, \ldots, 1, \ldots, L\}$ to form a discrete (Euler) approximation of Eq. 3.

We assume that the observation noise $e_m(t)$ is normally distributed with zero mean and variance ξ_m^2:

$$e_m(t) \sim N(0, \xi_m^2).$$

Note that the noise term $e_m(t)$ is uncorrelated across time points, and each ROI can have its own variance.

To control for computational burden in our applications below, $Y_m(t)$ is downsampled by a factor of 1,000 for each of $m = 1, \ldots, 6$. Specifically, we keep every 1,000th sample of $Y_m(t)$ and discard the others, a process that can be written as

$$\text{BOLD}_m(j) = Y_m(1000j),$$

where the index j is counted in seconds when applied to BOLD_m and in milliseconds when applied to Y_m. We perform this step to conform to the temporal resolution of fMRI BOLD signal in a real experiment, which depends on the repetition time (TR). We assume TR = 1s.

3. Model structure

Here, we apply the MDS framework to construct a generative model of perceptual decision making. In particular, we apply the MDS framework on a sequence of (assumed) experimental trials. The latent neuronal activation determines both BOLD signal (via the observation equation in Eq. 2) and behavioral data. The trials are consecutive so that neuronal activation in the current trial affects the following trial. The model is intended to describe how key ROIs systematically activate through time across experimental trials.

We first review some important findings about the neural substrate of perceptual decision making, because they provide the theoretical underpinnings of the model. Next, we construct a basis set of ROIs based on the literature, and then define the mathematical structure that relates the neuronal activation among the regions through time. As a reference, Fig. 3 shows the overarching structure of the model, where we assume a set of six ROIs (R_1, ..., R_6). By specifying a particular structural relationship between these regions, we can simulate the model's activity in the context of a random dot motion task.

The random dot motion task is often used to investigate the neural and cognitive basis of perceptual decision making (Ball and Sekuler, 1982; Britten et al., 1992; Churchland et al., 2008; Forstmann et al., 2010, 2008; Ho et al., 2009; Niwa and Ditterich, 2008; Roitman and Shadlen, 2002; Salzman and Newsome, 1994; Shadlen and Newsome, 2001; van Maanen et al., 2011). The stimuli in this task consist of an array of moving dots, where some percentage of the dots are moving in a coherent direction. The percentage of dots moving coherently can be varied, and this percentage is often treated as an independent variable to quantify the task difficulty (e.g., Britten et al., 1992).

The gist of how the brain processes information in the task can be described in three steps. First, sensory visual neurons in the brain areas MT and medial superior temporal (MST) of extrastriate cortex extract motion information from the visual image and represent the information within the visual cortex (Britten et al., 1992, 1996; Celebrini and Newsome, 1995; Croner and Albright, 1999; Shadlen et al., 1996). Neurons in MT and MST respond selectively to visual stimuli moving in particular directions reflecting the amount of motion energy to which they are tuned (Albright, 1984; Simoncelli and Heeger, 1998; Zeki, 1974). Second, the motion-direction representations in MT and MST are used to produce an integrated estimate of the net direction

of motion. There is evidence that the latter computation may be carried out in the frontal eye field (FEF) and the lateral intraparietal area (LIP) of the inferior parietal lobe (Andersen et al., 1992; Colby and Goldberg, 1999; Schall et al., 1995; Shadlen and Newsome, 2001). In particular, movement neurons in FEF and LIP initiate a saccade when their spike rate reaches a threshold (Brown et al., 2008; Dorris et al., 1997; Ratcliff et al., 2003, 2007). The cumulative strength of the motion information through time is often taken as evidence of accumulator dynamics in extant decision making models that assume sequential sampling of motion information (Boucher et al., 2007; Carpenter, 1999; Carpenter et al., 2009; Carpenter and Williams, 1995; Gold and Shadlen, 2007; Purcell et al., 2010; Ratcliff et al., 2003, 2007; Shadlen and Newsome, 2001).

The first and second steps only indicate the probability of making a decision choice to a certain direction for a given visual input, but the overt response relies on the ability of downstream neurons to select one unambiguous motor program and pass it on to the motor system for execution (Gold and Shadlen, 2001, 2002). This selection is thought to be performed by the superior colliculus (SC) and basal ganglia in the third step (Ding and Gold, 2013; Lo and Wang, 2006; Redgrave et al., 1999). The basal ganglia are known to have a critical role in voluntary motor behavior in general (Graybiel, 1995; Hikosaka et al., 2000b; Houk et al., 1995; Wickens, 1997). Neurons in substantia nigra pars reticulata (SNr), an output structure of the basal ganglia, send GABAergic projections to principal cells in the SC, providing a "default" level of tonic inhibition to the SC. This tonic inhibition is released when the SNr receives increased inhibitory inputs from caudate nucleus (CD, part of the striatum), which is driven by excitatory inputs from many cortical areas including the LIP and FEF (Hikosaka et al., 2000b, 2006). The third step of how LIP and FEF affect SC through the mediation of the basal ganglia is explained as a trade-off mechanism in the striatal hypothesis (Bogacz et al., 2010; Forstmann et al., 2008, 2010). The striatal hypothesis posits that an emphasis on speed promotes excitatory input from cortex to striatum; the increased baseline activation of the striatum acts to decrease the inhibitory control that the output nuclei of the basal ganglia exert over the brain, thereby facilitating faster but possibly premature responses.

Inspired by the aforementioned neural findings, in Fig. 3, R_1 and R_2 include visual neuronal populations mostly including MT and MST that selectively encode the motion information of the stimulus (Britten et al., 1992, 1996; Celebrini and Newsome, 1995; Croner and Albright, 1999; Shadlen

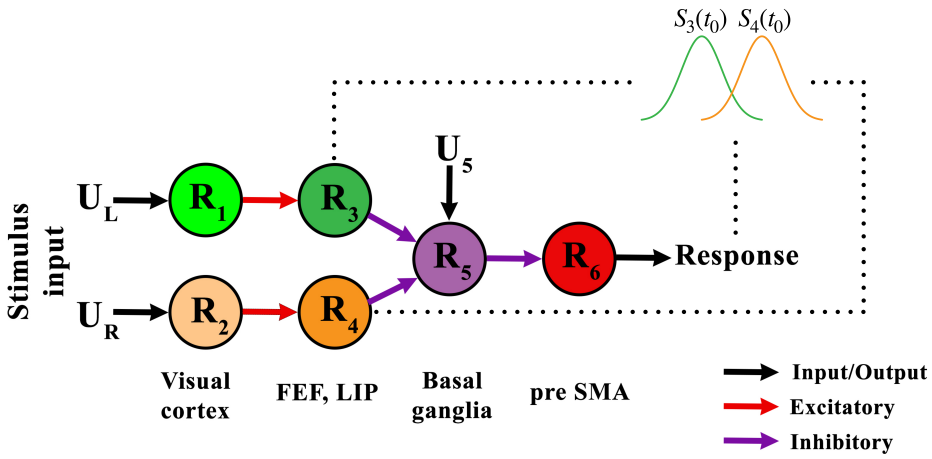

Figure 3: **A proposed MDS model for explaining neural and behavioral data from a perceptual decision making task.** U_L and U_R represent the visual inputs for leftward and rightward moving dots, respectively. R_1 and R_2 contain visual cortex neurons with direction-selective property. R_1 and R_2 encode visual inputs of random moving dots. R_3 and R_4, such as FEF and LIP, contain neurons that accumulate evidence for leftward and rightward dots independently. R_5 stands for the output nuclei of basal ganglia and R_6 represents pre SMA. U_5 provides a hypothetical constant input for R_5. Black arrows indicate input/output connections, red arrows indicate excitatory connections, and purple arrows indicate inhibitory connections. R_1 excites R_3, R_2 excites R_4, and R_5 inhibits R_6. When absolute difference of accumulated evidences between R_3 and R_4 reaches a threshold, R_3 and R_4 jointly inhibit R_5 so that R_6 gets disinhibited. R_6 accumulates evidence for the response options, eventually passing a signal to initiate a movement. The dotted lines represent a process of comparing values of $S_3(t_0)$ and $S_4(t_0)$ to determine the movement direction.

et al., 1996). The neuronal populations in R_1 are mainly sensitive to the leftward motion, whereas those in R_2 are mainly sensitive to the rightward motion. The direction-selective voxels in R_1 and R_2 can be decoded using multivoxel pattern analysis (MVPA) methods and fMRI (Kamitani and Tong, 2005, 2006; Serences and Boynton, 2007a,b). U_L and U_R in Fig. 3 are the leftward moving and rightward moving stimulus strengths for the nodes R_1 and R_2, respectively.

R_3 and R_4 contain neuronal populations of FEF and LIP that further process the visual information from R_1 and R_2 to guide the responses. Hence, the neuronal activations in R_1 and R_2 induce the neuronal activations in R_3 and R_4 respectively, through their endogenous connectivity, and this induction is illustrated as red arrows in Fig. 3 to represent excitatory effects. Here we conceptualize the instantaneous neuronal activations in R_3 and R_4 as two independent decision variables evolving at each time point, which makes R_3 and R_4 function as two independent accumulators. This independent accumulator assumption has been used in many perceptual decision-making models (Boucher et al., 2007; Carpenter, 1999; Carpenter et al., 2009; Carpenter and Williams, 1995; Gold and Shadlen, 2007; Kim and Shadlen, 1999; Purcell et al., 2010; Ratcliff et al., 2003, 2007; Schall, 2003; Shadlen and Newsome, 2001).

Moving rightward along the diagram in Fig. 3, R_5 is assumed to be the output nuclei of basal ganglia and R_6 is assumed to be the presupplementary motor area (pre SMA). The neuronal activation in R_5 continuously sends tonic inhibition to R_6, preventing R_6 from making a response, and this tonic inhibition is illustrated as a purple arrow from R_5 to R_6 in Fig. 3. U_5 provides a hypothetical constant impulse input for R_5 so that when there is no other brain region connected with R_5 and R_6, R_5 remains positively activated and thus R_6 remains inhibited. Although there are many other regions (e.g. SC, striatum) that play an important role in decision making, we have omitted these areas from the MDS model for simplicity and their activities are unlikely to be clearly measured in real experimental data.

R_3 and R_4 are conditionally connected with R_5 through a dynamic gating mechanism. The dynamic gating mechanism has been widely adopted to explain how interactions between basal ganglia and cortical regions affect information updating inside the cortical regions (O'Reilly, 2006; Redgrave et al., 1999; Stewart et al., 2010). Following the notation of MDS, we express $E(t)$ as

$$E(t) =\mid S_3(t) - S_4(t) \mid,$$

where $S_3(t)$ and $S_4(t)$ indicate the neuronal activation in R_3 and R_4 at time point t, respectively. Whenever $E(t)$ reaches a pre-specified threshold value θ_1, the connections from R_3 and R_4 to R_5 are initiated, illustrated as purple arrows from R_3 and R_4 to R_5 to represent inhibitory effects. We denote the time at which the threshold is reached as t_0. The relative values of $S_3(t)$ and $S_4(t)$ determine the response: a leftward choice is made if $S_3(t_0) > S_4(t_0)$, and a rightward choice is made otherwise. The values of $S_3(t)$ and $S_4(t)$ are each depicted by a Gaussian distribution in Fig. 3, where in this illustration rightward motion (U_R) is stronger than leftward motion (U_L) on average. The joint inhibition from R_3 and R_4 makes R_5 unable to inhibit R_6 (Bogacz et al., 2010; Forstmann et al., 2008, 2010; Hikosaka et al., 2000b, 2006). Notice that the involvement of dynamic gating mechanism changes the connectivity matrix after t_0. This is the main reason why $C(t)$ in Eq. 1 is time-variant.

Once R_6 becomes disinhibited, the neuronal activation in R_6 is monitored and accumulated at each moment from t_0. As soon as it reaches a pre-specified threshold θ_2, R_6 sends out a signal to the muscle to initiate a movement, denoted as t_1 (Forstmann et al., 2008; Georgiev et al., 2016; Mansfield et al., 2011). The response to be made is determined by the relative magnitude of neuronal activations in R_3 and R_4 at t_0. As modeling motor control is beyond our present scope, we assume a constant delay parameter τ to execute the movement. This parameter is often used to model non-decision processes in other decision making models (Brown and Heathcote, 2005, 2008; Ratcliff and Smith, 2004; Smith and Vickers, 1988).

Following a response, the visual inputs from external stimuli are switched off (i.e. the values of U_L and U_R return to zero). As a result, the mean activations of R_1, R_2, R_3 and R_4 return to zero, but fluctuate around this mean due to the noise term in Eq. 1. The variation of the noise term in R_1 and R_2 decreases after making a response to represent the lower noise variations of neuronal activations in visual processing ROIs after the response being made. Meanwhile, the joint inhibition from R_3 and R_4 to R_5 is cancelled. R_5 becomes disinhibited and R_6 reverts back to being inhibited. The system remains at this stasis point until another stimulus is presented.

Fig. 4 illustrates an example of how this model works by showing a trial of latent brain activity evolution of the six ROIs in Fig. 3. The three panels

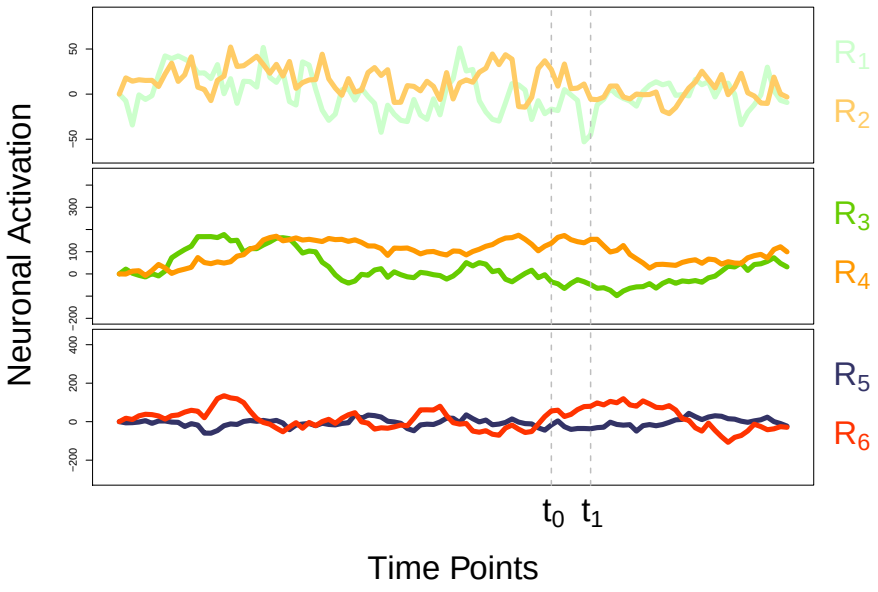

Figure 4: **An example showing the states of neuronal activation corresponding to six hypothetical regions of interest (ROIs) in Fig. 3.** The x-axis represents the time course of a trial and the y-axis represents the neuronal activation. Each line is associated with a ROI. The time point t_0 is when the absolute difference of accumulated evidences between R_3 and R_4 reaches a predefined threshold. The time point t_1 is when R_6 is ready to initiate a response.

show how neuronal activations (y-axis) evolve with time (x-axis) in the six ROIs (R_1 to R_6). ROIs are colored corresponding to the nodes in Fig. 3. U_R is set to be 4 times larger than U_L. By observing the neuronal activation oscillations of R_1 and R_2 in the top panel before t_1, there is a clear pattern that the magnitude of activation within R_2 is higher than that of R_1. The opposite pattern (i.e., R_1 is higher than R_2) mainly arises from the large noise term. We use the same connectivity coefficient from R_1 to R_3 and from R_2 to R_4. As such, in the middle panel, most of the time R_4 lies above R_3 before time point t_1, similar to the pattern in the top panel where R_2 lies above R_1. The bottom panel shows the effect on R_6 of the tonic inhibition from R_5 for the time points before t_0. Here, the activations of R_5 and R_6 are interwoven with each other whereas R_5 looks more stable. The bottom panel within the time window from t_0 to t_1 illustrates a different pattern of activations between the R_5 and R_6 nodes. While R_6 rises rapidly and remains highly activated, R_5 remains negatively activated. The rightmost portion of the graph after the movement-initiation time t_1 illustrates the neuronal activations in the six ROIs after making the response. The activations of R_1, R_2, R_3 and R_4 fluctuate around zero means. R_5 becomes disinhibited and R_6 is inhibited at the negative value.

The MDS model we have developed can be used to generate predictions about neural and behavioral data through simulations. Although the model has several components and temporally-specific changes to its parameters, Appendix A provides pseudocode with explicit steps detailing these changes to facilitate the model's implementation. As a test of the model's appropriateness, in the following sections, we simulate the model under different stimulus configurations, and evaluate whether the model generates patterns of data that are reasonable. In Simulation Study 1, the stimulus strength (coherence) favored either one or the other response on most trials, which is a typical experimental procedure. In Simulation Study 2, the two kinds of visual motion were balanced on all trials while their (common) absolute coherence was manipulated. While such balanced coherencies for the two options should not present a problem in principle, this particular stimulus configuration presents an interesting challenge to many extant models of decision making (Ratcliff et al., 2018; Teodorescu and Usher, 2013; Teodorescu et al., 2016).

4. Simulation Study 1: Unequal coherence

The first simulation involves a standard set of stimuli, where coherence is varied along a single dimension, varying in strength for leftward and rightward response options. For the purposes of our simulation, we assumed 1,000 dots shown on the screen, with each one moving either leftward or rightward. Then the leftward dots and rightward dots can be subtracted from each other to form a *net* coherence. For example, if there are 30% leftward dots and 70% rightward dots, then the net coherence level is 40% to the right. The probability of leftward dots is defined as p_L and it is the independent variable in this simulated experiment. Table 1 shows that p_L varies from .1 to .9, increasing by .1, implying that the probability of rightward dots p_R decreases from .9 to .1 by .1. The net coherence equals to the absolute difference of p_L and p_R, with the direction determined by the larger one of p_L and p_R.

p_L	.1	.2	.3	.4	.5	.6	.7	.8	.9
p_R	.9	.8	.7	.6	.5	.4	.3	.2	.1
Net coherence	.8	.6	.4	.2	0	.2	.4	.6	.8

Table 1: The p_L condition levels, corresponding p_R levels and net coherence levels used in Simulation Study 1. p_L: probability of leftward moving dots. p_R: probability of rightward moving dots. The net coherence is the absolute difference of p_L and p_R, with the direction determined by the larger one of p_L and p_R.

With p_L and p_R at hand, we can calculate the number of leftward dots and rightward dots and use them to represent the strengths of the visual stimuli. In the simulation, we use the number of dots as strength of input of U_L and U_R. For each time point t from stimulus onset to movement-initiation time t_1, the number of leftward moving dots (U_L) is randomly sampled from a Binomial distribution with a given probability parameter p_L:

$$U_L(t) \sim \text{Binomial}(1000, p_L), t = 1, \ldots, t_1$$

and the number of rightward moving dots (U_R) equals to U_L subtracted from 1,000:

$$U_R(t) = 1000 - U_L(t), t = 1, \ldots, t_1.$$

Hence, as U_L is sampled at each moment in time, the stimulus is stochastic, and the strength of evidence fluctuates through time. The sum of $U_L(t)$ and

$U_R(t)$ is always a fixed 1,000. Following a response (i.e. $t > t_1$), $U_L(t)$ and $U_R(t)$ are set to zero.

$U_L(t)$ and $U_R(t)$ are used as first two elements in the external input vector $U(t)$ in Eq. 1, making U_L and U_R the impulse functions for R_1 and R_2, respectively. The values of $U_L(t)$ and $U_R(t)$ are both divided by 100 to scale the strength of neuronal activation. U_5 is fixed to be 1 and passed to the fifth element in the vector $U(t)$, implying a hypothetical constant magnitude of impulse function for R_5.

We simulated a series of 270 trials where each trial is associated with a p_L condition. We assumed 30 trials for each p_L condition, and different p_L conditions are randomly interleaved across trials. Hence, this simulated experiment can be considered as an event-related design. Each series with 270 trials can be simulated multiple times to take into account the randomness in a simulated experiment. We simulated the series of trials for 100 times in order to observe the data pattern, and the randomization of trial conditions is fixed during the replication.

4.1. Parameters

The state equation (Eq. 1) and observation equation (Eq. 2) in MDS and perceptual decision making model structure in Section 3 involve many parameters. In this section, we describe how we specified those parameters for the current simulation. We adopt two different forms for the intrinsic connectivity matrix $C(t)$ according to the model structure in Section 3, hence the time-dependent specification of the C matrix stated earlier. Specifically, we let $C(t) = C_1$ from the beginning of a trial until the threshold-crossing time ($t < t_0$) and then again from the motion-initiation time ($t > t_1$) until the end of the trial, where

$$C_1 = \begin{pmatrix} .5 & 0 & 0 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 & 0 & 0 \\ .7 & 0 & .9 & 0 & 0 & 0 \\ 0 & .7 & 0 & .9 & 0 & 0 \\ 0 & 0 & 0 & 0 & .7 & 0 \\ 0 & 0 & 0 & 0 & -.8 & .7 \end{pmatrix}.$$

The diagonal elements in matrix C_1 indicate that the within-region connectivity strengths are .5 in R_1 and R_2, .9 in R_3 and R_4, and .7 in R_5 and R_6. The self connectivity strengths were all set to be .7 in Ryali et al. (2011)

and we used this value for R_5 and R_6. The instantaneous neuronal activations in R_3 and R_4 are assumed as a result of accumulated evidence so their self connectivity strengths have to be larger and close to 1. R_1 and R_2 process visual stimuli so their self connectivity should be smaller than R_3 and R_4. We used .5 to allow some amount of leakage to represent the mechanism that part of visual stimulus information is lost in visual stimulus processing (McClelland, 1993; Smith, 1995; Usher and McClelland, 2001). Then $C_1[3,1] = C_1[4,2] = .7$ indicates that the connectivity strengths from R_1 to R_3 and from R_2 to R_4 are both .7. This medium high value indicates the proportion of information is passed by from R_1 and R_2 to R_3 and R_4, respectively at each moment. Note that we assume a symmetric pattern in the leftward motion pathway ($C_1[1,1]$, $C_1[3,3]$ and $C_1[3,1]$) and rightward motion pathway ($C_1[2,2]$, $C_1[4,4]$ and $C_1[4,2]$) by equal connectivity strengths. $C_1[6,5] = -.8$ indicates a negative connectivity strength -.8 from R_5 to R_6, and this negative connectivity represents the constant inhibition from R_5 to R_6. All the other connectivity strengths were set to zero.

During the interval from the threshold-crossing time to the motion-initiation time ($t_0 \leq t \leq t_1$), the connectivity matrix changes to $C(t) = C_2$, where the matrix C_2 is identical to C_1 except that $C_2[5,3] = C_2[5,4] = -.2$. That is,

$$C_2 = \begin{pmatrix} .5 & 0 & 0 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 & 0 & 0 \\ .7 & 0 & .9 & 0 & 0 & 0 \\ 0 & .7 & 0 & .9 & 0 & 0 \\ 0 & 0 & -.2 & -.2 & .7 & 0 \\ 0 & 0 & 0 & 0 & -.8 & .7 \end{pmatrix}.$$

This change indicates that the connectivity strengths from R_3 to R_5 and from R_4 to R_5 are both -.2 after the connection paths are switched to "on" mode at threshold-crossing time t_0, and are later changed back to zero at t_1 when the responses from R_6 are initiated. The connection strengths are back to zero after t_1 to prepare for stimulus presentation in the next new trial. The change from C_1 to C_2 and back to C_1 is possible via the dynamic gating mechanism.

The direct input matrix D is composed with diagonal elements for those

regions with an external input (i.e. R_1, R_2 and R_5), so

$$D = \begin{pmatrix} .9 & 0 & 0 & 0 & 0 & 0 \\ 0 & .9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The direct effect coefficients of U_L, U_R and U_5 on R_1, R_2 and R_5 are all .9. This value was set to be slightly lower than 1 to represent the fact that the strength of physical stimuli (i.e. U_L, U_R) can only be partially captured by visual processing regions (i.e. R_1, R_2).

The noise term $\omega(t)$ was distributed according to a multivariate normal distribution in 6 dimensions, with zero mean and a diagonal variance-covariance matrix, which indicates uncorrelated noise across the 6 ROIs. Let $\sigma^{(m)}(t)$ denotes the standard deviation of the noise in the m-th ROI. We set all 6 standard deviations to the same value $\sigma_1 = 16$ throughout the time interval before initiating the motor response ($t < t_1$). After t_1, the noise in the two sensory ROIs was reduced to $\sigma^{(1)}(t) = \sigma^{(2)}(t) = \sigma_2 = 5$, whereas the noise in the other regions remained at its former level $\sigma_1 = 16$.

The parameters used for observation equation step in Eq. 2 were set as follows. The length L of the HRF function was 32s so there were 32,000 data points in each $h_m(\tau)$ in the temporal unit of 1 millisecond. The amplitude parameters for the 6 ROIs were $A_1 = A_2 = .0005$, $A_3 = A_4 = .00006$, $A_5 = .0015$, and $A_6 = .0002$. These A's scale the BOLD signal to be approximately within the range from -2 to 2. The amplitude parameters were the same for R_1 and R_2, and same for R_3 and R_4, so that the magnitudes of BOLD signals of R_1 and R_2, and of R_3 and R_4 were comparable. The standard deviation ξ_m of the observation error term was set to .05 for all ROIs. We performed the linear convolution in Eq. 2 in frequency domain for fast computation using a C subroutine library FFTW 3.3.8 (Frigo and Johnson, 2005).

In the model structure described in Section 3, the threshold θ_1 was set to be 250 and θ_2 was 1,500. The non-decision time τ was set to be 100. The number of total time points allowed in one trial was 2,000 (i.e. 2s).

We further simulated a rapid event-related fMRI design with 30 trials at each p_L condition. Conditions were interleaved to create a time series of trials. In the series of trials, the latent neuronal activation at the last time

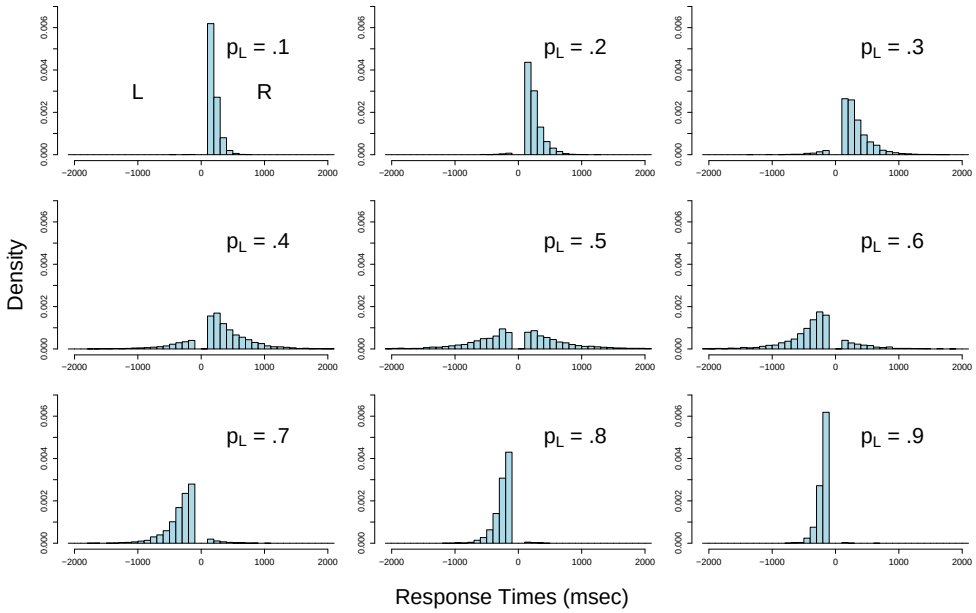

Figure 5: **Behavioral choice and response time distributions from Simulation 1.** Each histogram corresponds to a p_L level. Response times corresponding to the left choice (L) are shown on the negative $x-$axis, whereas response times corresponding to the right choice (R) are shown on the positive $x-$axis.

point of each trial for each ROI was used as the latent neuronal activation at the starting time point of the next trial for the corresponding ROI. This operation allows trial-to-trial dependencies in the time series data. Each time series can be simulated multiple times and so we set the number of simulations to be 100 to accurately reflect the patterns in the data. Therefore, each p_L condition was repeated for 3,000 times.

4.2. Results

Under the simulation setup, 99.84% of all the trials produce a left or right response within 2s. Fig. 5 shows the behavioral choice and response time data from Simulation 1. The nine panels are corresponding to the conditions of p_L from .1 to .9. In each panel, response times corresponding to the left choice (L) are shown on the negative $x-$axis, whereas response times corresponding to the right choice (R) are shown on the positive $x-$axis. This simulation result indicates that as p_L increases, the proportion of the

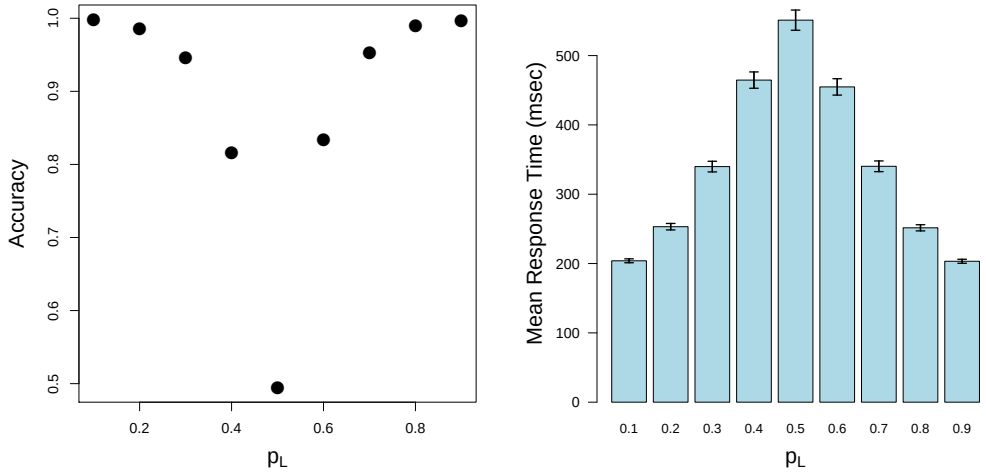

Figure 6: **Summary of behavioral data from Simulation 1.** The left panel shows choice accuracy at each p_L level where accuracy is defined as correctly choosing the direction with more moving dots. When p_L is .5, choosing right direction is defined as the correct choice. The right panel shows the mean response time at each p_L level. Error bars are included considering the number of simulation (3,000) at each p_L level (excluding number of non-response trials), and are extended to ± 2 standard errors about the mean response times.

left choice increases, along with the decrease of the right choice. When p_L equals to .5, approximately the same number of choices are made between left and right alternatives (49.48% of right choice in the simulation). Recall that p_L represents the input strength of the leftward motion relative to the total input strength of leftward and rightward motion. Therefore, when p_L increases, the input strength for the leftward motion increases, along with the decrease of the input strength for the rightward motion, so the proportion of the leftward choices increases.

We summarize the simulated behavioral results in Fig. 6 by showing how accuracy and mean response time change with p_L. In the left panel, accuracy is defined as correctly choosing the direction with more moving dots. Fig. 6 shows that accuracy decreases when p_L increases from .1 to .5 and increases when p_L goes from .5 to .9. The accuracy pattern is symmetric around $p_L = .5$, and this symmetric pattern is also shown on the mean response time in the right panel. The symmetric pattern in the behavioral data originates from the symmetric net coherence in Table 1. Previous studies with the

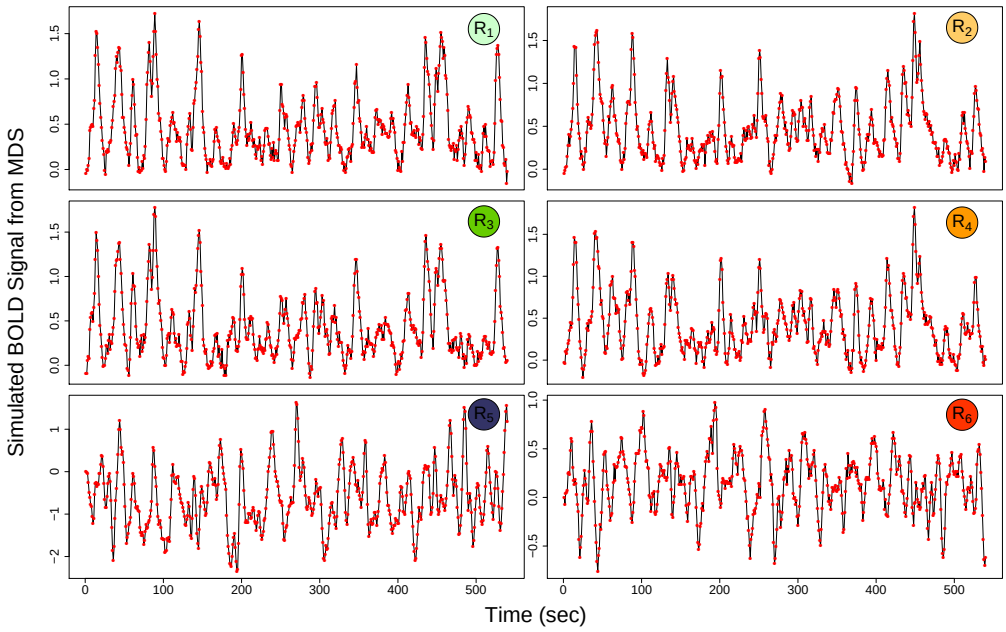

Figure 7: **Simulated BOLD signal for each ROI from Simulation 1.** The red dots are signals after downsampling and black lines connect neighboring dots.

random dot motion paradigm have shown that the net coherence serves as an indicator for the difficulty of the task (Britten et al., 1992; Roitman and Shadlen, 2002; Salzman and Newsome, 1994; Shadlen and Newsome, 2001). Therefore, our simulated random dot motion task becomes harder as p_L goes from .1 to .5 and becomes easier as p_L goes from .5 to .9, which is reflected in the behavioral data.

Fig. 7 illustrates an example of the simulated BOLD signal for six ROIs from one simulation. The red dots are signals after downsampling and black lines connect neighboring dots. Every trial has a duration of 2s and there are 270 trials in one simulation. With assumed TR = 1s, we have 2 samples in each trial and thus 540 BOLD data points in total. The shape of the oscillations is due to the HRF shape in Eq. 3 and is similar with the typical BOLD signal from real experiments. We observe that there are very similar simulated BOLD signals in R_1 and R_3, and also in R_2 and R_4, across trials. This similarity is expected because in every trial, the neuronal activations in R_3 and R_4 are directly affected by R_1 and R_2, respectively. The differences across time points between R_1 and R_3 and between R_2 and R_4 within each

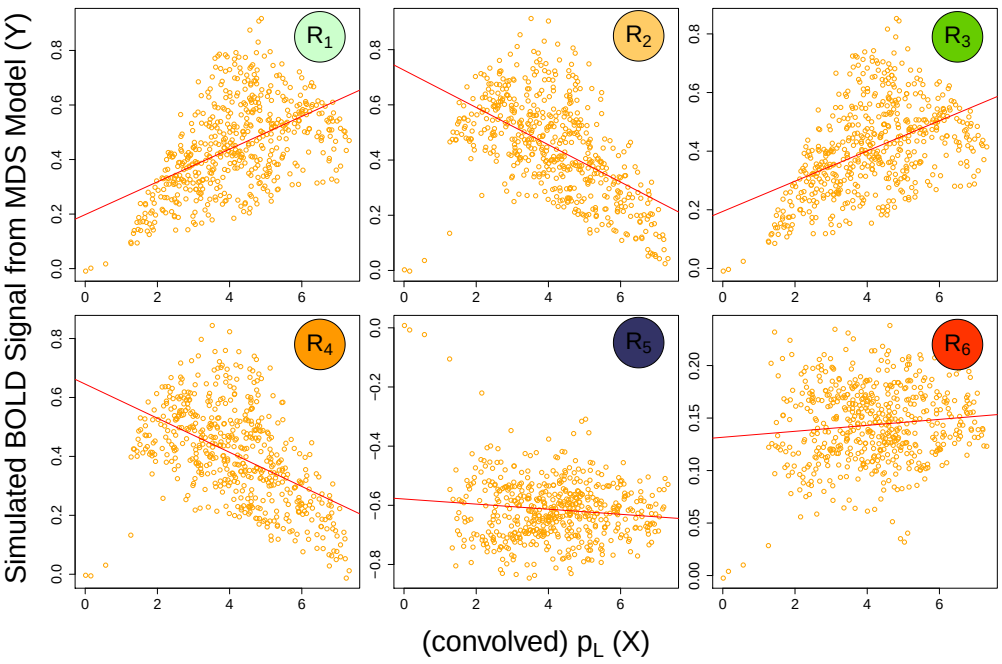

Figure 8: **GLM fits for simulated BOLD signal from MDS model in Simulation 1 and the convolved p_L conditions in each ROI.** The orange dots illustrate the correlational patterns in each ROI and red straight lines show least square fits of GLMs.

trial (as illustrated in Fig. 4) are cancelled in the downsampling process.

To further investigate the pattern of neural data, we fit the general linear model (GLM) to the simulated BOLD signal in each ROI, with the interleaved 9 p_L conditions (.1, .2, ..., .9) as the explanatory variable. Following traditional fMRI data analysis procedure, we convolved the interleaved 9 conditions with the same HRF in Eq. 3 before fitting the GLM and denote convolved p_L conditions as $\mathbf{X}$. For the simulated BOLD signal, we computed the average of BOLD signal over 100 simulations for six ROIs and denote as a matrix $\mathbf{Y}$. GLM assumes

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E},$$

where β is the linear coefficient vector and $\mathbf{E}$ is an uncorrelated error term following the multivariate normal distribution. Fig. 8 illustrates correlational pattern in each ROI with the orange dots and shows the least square fit of GLM by the red straight lines. There are significant correlations in R_1 ($\hat{\beta}_1 = .044, p = 1.48\text{e}{-11}$), R_2 ($\hat{\beta}_2 = -.057, p = 2.77\text{e}{-16}$), R_3 ($\hat{\beta}_3 = .038, p = 9.56\text{e}{-10}$), and R_4 ($\hat{\beta}_4 = -.050, p = 1.37\text{e}{-14}$), but are not significant correlations in R_5 and R_6. This pattern is consistent with the model assumption, as R_1 and R_2 process visual motion for leftward and rightward moving, respectively, and R_3 and R_4 accumulate neuronal evidence of leftward and rightward motion, respectively. Although the pair of R_1 and R_3, and the pair of R_2 and R_4 both produce similar BOLD signals as shown in Fig. 7, the correlations in R_1 and R_2 are slightly stronger than those in R_3 and R_4, respectively. R_1 and R_2 process the input of motion information (i.e. U_L and U_R) directly, but R_3 and R_4 access the motion information indirectly mediated by R_1 and R_2. More motion information is lost after passing through R_1 and R_2.

As a short conclusion, Simulation 1 generated behavioral and neural data from a random dot motion task with unequal coherence. Both behavioral and neural data show qualitatively comparable characteristics with real experimental data.

5. Simulation Study 2: Balanced coherence

In Simulation 1, we treated the dots moving towards other directions other than leftward or rightward as the irrelevant "noise" and did not explicitly

model them. Although behavioral and neural predictions in Simulation 1 were qualitatively similar to real data, it is worth considering if these dots can be treated as irrelevant noise. In the next simulation, we aim to simulate the scenario where the leftward dots and rightward dots are of equal amount but the ratio of their summed amount compared to the total amount of dots vary. This way, we are able to detect if the moving dots towards other directions play a role in the random dot motion paradigm. If these dots are indeed irrelevant, the model should predict similar predicted behavioral and neural results with the varied ratio. However, if these dots have an effect on the decision making process, the predicted behavioral and neural data could provide insights to the understanding of the problem.

In fact, this situation of balanced coherence is related to the argument between sensitivity to absolute values and to relative values (Teodorescu et al., 2016). Promoters of the relative values postulate that decision making is guided by the relative value difference of the two alternatives, in terms of either the difference or the ratio (Brown and Heathcote, 2008; Ratcliff and Rouder, 1998; Roe et al., 2001; Tversky and Simonson, 1993). However, others argue that task irrelevant absolute values are also important (Usher and McClelland, 2001). In other words, the absolute value of the alternatives cannot be simply represented by the relative value. Behavioral data from equal-valued decision making tasks show that equal-but-low-value alternatives need longer processing time compared with equal-but-high-value alternatives (Pirrone et al., 2014; Teodorescu et al., 2016), implying the importance of absolute value of choice alternatives. The balanced coherence of leftward and rightward dots thus provides an appropriate emulation of this situation.

5.1. Parameters

We still used 1,000 random dots in total, but the 1,000 random dots contained the same probability of leftward and rightward dots, along with dots in other arbitrary moving directions. The effects of dots in other moving directions were still assumed to be offset by summing up. In the 1,000 dots, the probability of leftward dots p_L (and also rightward dots p_R) was manipulated across .1, .2, .3, .4 and .5, so the probability of the dots moving towards other directions (p_{others}) was correspondingly .8, .6, .4, .2 and 0. From stimulus onset to movement-initiation time t_1, the numbers of leftward dots, rightward dots and the others were randomly sampled from a Multinomial distribution, such that

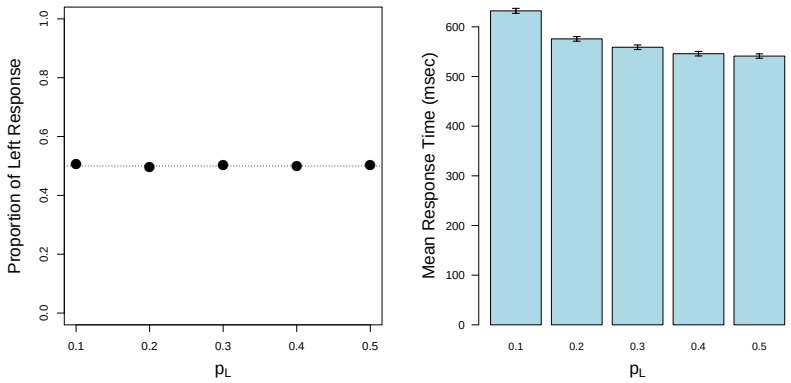

Figure 9: **Summary of behavioral data from simulation 2.** The left panel shows proportion of left response at each p_L level where a dashed line indicates proportion of .5 as a reference. The right panel shows mean response time at each p_L level using the barplot. Error bars are included considering the number of simulation (3,000) at each p_L level (excluding number of non-response trials), and are extended to ± 2 standard errors about the mean response times.

$$(U_L(t), U_R(t), U_{\text{others}}(t)) \sim \text{Multinomial}(1000, (p_L, p_R, p_{\text{others}})), t = 1, \ldots, t_1$$

where U_L, U_R and U_{others} are the number of dots moving towards left, right and any other direction, respectively. p_L, p_R and p_{others} are Multinomial distribution parameters indicating the probabilities for random dots to move towards each of the directions. $U_L(t)$ and $U_R(t)$ were again both divided by 100. The other parameter settings in this simulation were exactly the same as those in Simulation 1.

5.2. Results

Under the simulation setup, 99.12% of all the trials produce a left or right response within 2s. Fig. 9 shows the simulated behavioral data of response proportion and mean response time as p_L (or p_R) increases from .1 to .5, with non-response trials excluded. Not surprisingly, almost equal proportion of left choices and right choices are made across p_L (or p_R) from .1 to .5, because the input stimuli provide equal amount of strength for the leftward and rightward moving dots. However, the mean response time shows a decreasing trend as p_L increases. This result successfully recovers

the key response time data findings in previous studies (Pirrone et al., 2014; Teodorescu et al., 2016).

To investigate the pattern of neural data, we fit GLM to the simulated BOLD signal in each ROI, with the interleaved 5 p_L conditions (.1, .2, . . . , .5) as the explanatory variable. Fig. 10 shows correlational patterns in each ROI, same as in Fig. 7. The least square fits in red straight lines give significant correlations in all six ROIs: R_1 ($\hat{\beta}_1 = .173, p = 2.2\text{e}{-16}$), R_2 ($\hat{\beta}_2 = .171, p = 2.2\text{e}{-16}$), R_3 ($\hat{\beta}_3 = .164, p = 2.2\text{e}{-16}$), R_4 ($\hat{\beta}_4 = .160, p = 2.2\text{e}{-16}$), R_5 ($\hat{\beta}_5 = -.058, p = 3.7\text{e}{-9}$), and R_6 ($\hat{\beta}_6 = .009, p = .008$). Comparing with Fig. 7, the correlations are much stronger in all six ROIs, and correlations in R_5 and R_6 are both significant in terms of p-value. However, the plots of R_5 and R_6 do not show any correlational patterns. The significant correlations are likely to be driven by a few potential "outlier" points.

As a short conclusion, Simulation 2 adopted balanced coherence for two directions and generated behavioral choice and response time data, consistent with empirical findings. This simulation endorses the ability of the MDS framework to accommodate different type of task configuration in simulating behavioral and neural data.

6. Fitting the Model to Data

In this section, we investigate the model's inferential properties, with three questions in mind. First, can the model be fit to data? For models like MDS, this is a complicated problem as the model must be fit to the entire time series of neural and behavioral data. Importantly, as we do not assume that data are independent and identically distributed, fitting the model to data also entails capturing trial-to-trial dependencies (Turner et al., 2015b; Turner, 2019; Wagenmakers et al., 2004). Second, are the model parameters identifiable? Identifiability refers to a property of a model such that any particular parameter value maps to a unique probability density function (Bamber and Van Santen, 2000). Hence, our goal is to provide some initial evidence that for a given distribution of data, the model parameters have a unique solution. Third, if successful in fitting the model to data, are the recovered parameters veridical? In other words, are the estimated parameters similar to the true parameters used to generate the data?

To investigate these questions, we performed a model recovery study. We first generated synthetic data – similar to the experiments reported in the previous two sections – and then fit the model to the generated data. We

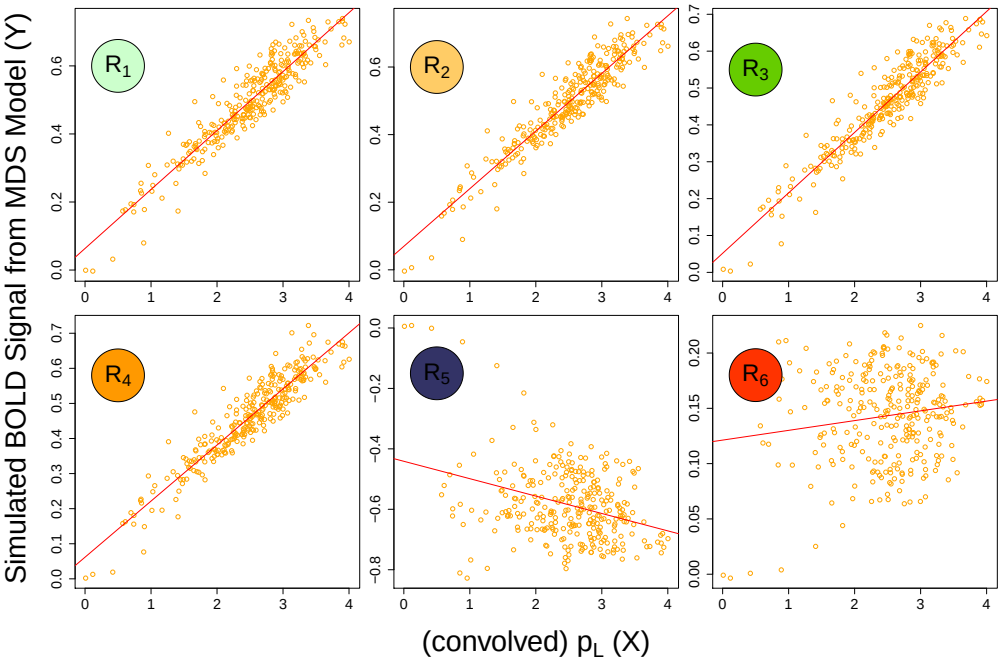

Figure 10: **GLM fits for simulated BOLD signal from MDS model in Simulation 2 and the convolved p_L conditions in each ROI.** The orange dots illustrate the correlational patterns in each ROI and red straight lines show least square fits of GLMs.

chose to use Bayesian inferential techniques because the resulting posterior distribution would allow us to simultaneously assess both the accuracy of the estimates (i.e., the central tendency of the posterior), and the uncertainty about them (i.e., the posterior's spread over the parameter space). As the model's likelihood function is analytically intractable, we used approximate Bayesian methods to form an approximation of the likelihood. Hence, our statistical methods enable us to answer our second question by assessing the shape (i.e., the concavity) of the posterior distribution, and our third question by comparing the central tendency of the posterior to the true values used to generate the data. The first question is answered by procuring solutions to the latter two questions.

6.1. Data Generation and Problem Statement

To emulate a real experimental setting, we assumed that each of the nine p_L conditions consisted of 30 trials, and all the conditions were interleaved. Therefore, the generated data consists of 270 trials, where each trial is associated with a choice and response time, except for one trial where no response is made during 2s. The choice for the non-response trial is coded differently than the other trials and its existence shows no effect on the estimation process. In addition, each trial period has an associated neural time series matrix where we assumed a one second fMRI acquisition sequence (i.e., the TR). Because each stimulus presentation period lasted for two seconds, each of the six neural time series vectors consisted of 540 data points. Notice that each time series was simulated for just one time, different from Section 4 and Section 5 where each time series was simulated for 100 times. Hence we would reasonably expect this randomness in data generation to be included in the posterior estimates.

Estimating the full matrices C_1, C_2 and D in Eq. 1 poses a great computational challenge. We chose to limit our scope by decomposing these matrices into their key individual elements, where Table 2 lists all of the important parameters from this decomposition. The first column provides the parameter notation, the second column describes the parameter's function, and the third column is the true value that was used to generate the dataset. For this analysis, we focused on recovering five key parameters: a_1, c_2, θ_1, σ_1 and A_{12}. We chose parameters a_1, c_2 and σ_1 from MDS state equation (Eq. 1) , A_{12} from observation equation (Eq. 2), and θ_1 from the model structure, trying to include parameters from different sources.

For the purposes of recovery, we allowed these five parameters to freely vary while keeping other parameters in Table 2 fixed to their true values. In terms of implementation, estimation requires that we search the space of all possible combinations of the model parameters, and evaluate their relative probabilities of having generated the data.

Parameter	Description	value
c_0	Within-region connection strength of R_5 and R_6	.7
c_1	Within-region connection strength of R_1 and R_2	.5
c_2	Within-region connection strength of R_3 and R_4	.9
a_1	Connection strength from R_1 to R_3 and from R_2 to R_4	.8
a_2	Connection strength from R_3 to R_5 and from R_4 to R_5	-.2
a_3	Connection strength from R_5 to R_6	-.8
d_1	Direct effect from U_1 to R_1 and from U_2 to R_2	.9
d_2	Direct effect from U_5 to R_5	.9
τ	Non-decision time	100
θ_1	Threshold value for the difference between R_3 and R_4	250
θ_2	Threshold value for accumulated movement information in R_6	1,500
σ_1	Standard deviation of the noise term before t_1	16
σ_2	Standard deviation of the noise term in R_1 and R_2 after t_1	5
A_{12}	Magnitude parameter in the canonical HRF function for R_1 and R_2	.0005
A_{34}	Magnitude parameter in the canonical HRF function for R_3 and R_4	.00006
A_5	Magnitude parameter in the canonical HRF function for R_5	.0015
A_6	Magnitude parameter in the canonical HRF function for R_6	.0002
ξ_m	Standard deviation of observation error of BOLD signal	.05

Table 2: **Summary of parameter in the perceptual decision making MDS model.** The first column provides the parameter notation, the second column describes the parameter's function, and the third column is the true value that was used to generate the dataset.

6.2. Estimation Methods

When using Bayesian statistics, acquiring any posterior distribution depends on efficient evaluation of two functions: (1) the prior distribution for the model parameters, and (2) the likelihood function relating the model parameters to the observed data. The posterior distributions $\pi(\theta \mid X^O)$ reflect our knowledge about a parameter set θ after observing a dataset X^O, and it is obtained by combining the prior $\pi(\theta)$ with the likelihood of a parameter set θ:

$$\pi(\theta \mid X^O) \propto \pi(\theta) L(\theta \mid X^O). \tag{4}$$

The prior distribution $\pi(\theta)$ reflects our knowledge of the parameter set θ before observing X^O, and is typically easy to specify in terms of defining its functional form. However, the likelihood function $L(\theta \mid X^O)$ is often much more difficult to derive. For simulation-based models that attempt to provide mechanistic explanations for how data manifest, direct evaluation of the likelihood function can be difficult, if not impossible. Unfortunately for us, the MDS model is one such simulation-based model with complex, stochastic characteristics, and these features of the model make its likelihood function intractable.

6.2.1. Likelihood Estimation: Kernel-Based ABC

To approximate the likelihood function of the MDS model, we used kernel-based approximate Bayesian computation (KABC) method (Palestro et al., 2018b; Turner and Sederberg, 2012; Turner and Van Zandt, 2014, 2018; Turner et al., 2013a). As in a typical ABC approach, KABC requires that we first define a discrepancy function $\rho(\cdot)$, and use it to compare the "distance" between the simulated data (X^S) and observed data (X^O), where the simulated data are generated by $X^S \sim \text{MDS}(\theta)$ for a given parameter vector $\theta = \{a_1, c_2, \theta_1, \sigma_1, A_{12}\}$. When using KABC, we filter these distances by applying a continuous weighting function $\psi(\cdot|\delta)$ to $\rho(\cdot)$ to determine how closely X^S matches X^O. The parameter δ serves as a tuning parameter that controls the resolution of the "closeness" between X^S and X^O. When $\psi(\cdot)$ obeys certain properties (e.g., symmetric, unimodal), the term $\psi(\rho(X^S, X^O)|\delta)$ increases as X^S becomes more similar to X^O. As an example, perhaps the most common choice for $\psi(\cdot|\delta)$ is a Gaussian distribution centered at zero with standard deviation equal to δ. In this example, as δ decreases, larger weights in $\psi(\cdot|\delta)$ will be obtained if $X^S \approx X^O$, but a larger penalty will

be applied when the simulated data X^S are different from X^O. Hence, the choice of δ is an important one for accurately comparing X^S to X^O, an issue we discuss below.

For a static set of simulated data X^S, we could then just find the set of parameter values θ that maximize $\psi(\cdot|\delta)$, a relatively straightforward optimization problem. However, there is often considerable variability in the model generation process, where even for a fixed parameter value θ, we can arrive at very different sets of X^S. Hence, we can think of the data generation process as detailing a joint distribution over candidate parameter values θ and random realization of simulated data X^S. As our goal is to estimate θ and we do not care about the variability in X^S, we can obtain a posterior estimate by integrating out the variability in X^S:

$$\pi(\theta \mid X^O) \propto \int \pi(\theta)\text{Model}(x \mid \theta)\psi(\rho(x, X^S) \mid \delta)dx^S, \tag{5}$$

where $\text{Model}(x \mid \theta)$ denotes the density of data produced by the model simulation.

While the argument above has been constructed assuming X^O and X^S are scalars, for our estimation problem, our data consist of two time series vectors – one for choice data $\mathbf{C}$ and response time data $\mathbf{RT}$ – and one time series matrix consisting of BOLD signal data $\mathbf{Y}$ for each of the six ROIs. Hence, $X^O = \{C^O, RT^O, Y^O\}$ and $X^S = \{C^S, RT^S, Y^S\}$. To compare X^O and X^S, we assumed these variables were conditionally independent, and factorized the likelihood approximation:

$$\begin{aligned} L\left(\theta|X^O\right) &\propto \psi\left[\rho\left(X^O, X^S\right)|\delta\right] \\ &= \prod_i \psi\left[\rho\left(RT_i^S - RT_i^O\right)|\delta_1\right] \prod_i \psi\left[\rho\left(C_i^S - C_i^O\right)|\delta_2\right] \\ &\times \prod_m \prod_k \psi\left[\rho\left(Y_m^S(k) - Y_m^O(k)\right)|\delta_3\right]. \end{aligned} \tag{6}$$

To stabilize the variability in the data generation process (Toni et al., 2009), for each parameter proposal, we simulated the model 10 times and averaged the data C^S, RT^S, and $Y_m^S(k)$. With a suitable likelihood approximation in hand, we can substitute Eq. 6 into Eq. 4, and estimate the posterior distribution $\pi(\theta \mid X^O)$.

6.2.2. Posterior Sampling

We used numerical Monte Carlo approximation techniques to estimate the joint posterior distribution. Specifically, we used differential evolution with Markov chain Monte Carlo (DE-MCMC; ter Braak, 2006; Turner et al., 2013c, 2015b) to draw samples from Eq. 4. We chose DE-MCMC as it has been shown to be a highly efficient sampling method relative to MCMC, especially when sampling from posterior distributions whose parameter dimensions are correlated (Turner et al., 2013c).

Although Eq. 6 suggests that the tuning parameter $\delta = \{\delta_1, \delta_2, \delta_3\}$ are fixed, it is difficult to specify these parameters in advance. As we suggested above, finding the best values for δ is a difficult problem with grave consequences regarding the variance of the posterior distribution. Because decreasing δ increases the accuracy of the estimated posterior, one may be tempted to simply set δ to zero. However, with decreases in δ come other computational problems. Specifically, decreases in δ make it difficult to obtain high-quality estimates because of the sharp gradient associated with $\psi(\cdot|\delta)$. If the variability in the data generation process is large relative to the width of $\psi(\cdot|\delta)$, the chains of the sampling algorithm will tend to "stick" in the posterior distribution and will not sample from the posterior effectively.

To balance these two opposing forces, we used the Approximate Bayesian Computation with Differential Evolution (ABCDE) (Turner and Sederberg, 2012) algorithm to implement DE-MCMC sampling within the KABC likelihood approximation. ABCDE is unique as it uses two "modes" of sampling: a "burn-in" mode and an "sample" mode. In burn-in mode, ABCDE uses a specific optimization rule for moving the chains of the algorithm into the region of the posterior with highest density. To do this, ABCDE proceeds by optimizing Eq. 5 with respect to both θ and δ simultaneously. After some number of iterations, the values of δ asymptote to values that are as small as possible, but still enable efficient sampling from the posterior distribution of θ. After this point is reached, the algorithm switched to the sample mode, where δ is set to their lowest value obtained during the burn-in phase, and only θ is estimated.

We ran the burn-in phase of the ABCDE algorithm with 24 chains for 2,000 iterations, optimizing with respect to both θ and δ. After this initial phase, we set each δ to their respective (rounded) mean values, where $\delta_1 = 100$, $\delta_2 = 1$ and $\delta_3 = .4$, respectively. Henceforth, we used the sample mode of ABCDE to obtain posterior estimates of only θ, running the algorithm

for an additional 3,000 iterations, but discarded the first 1,500 iterations as an additional burn-in period (i.e., to allow the chains to spread out into the posterior distribution). Hence, our posterior estimates are based on 36,000 samples. A migration step (see Turner and Sederberg, 2012; Turner et al., 2013c) was used during the second burn-in period with probability 0.2 for the first 400 iterations, after which time the migration step was terminated. We also used a purification step every 10 iterations to ensure that the chains were not stuck in spuriously high regions of the approximate posterior distribution (Holmes, 2015). Convergence was checked by visual inspection.

We also estimated posterior densities by behavioral-only data or neural-only data to compare with the density from the joint model. Specifically, in behavioral-only estimation, $X^O = \{C^O, RT^O\}$ and $X^S = \{C^S, RT^S\}$, and in neural-only estimation, $X^O = \{Y^O\}$ and $X^S = \{Y^S\}$. The estimated likelihoods were constructed in the similar way as in Eq. 6, but we reduced the multiples according to the elements in X^O and X^S. We used $\delta_1 = 100$ and $\delta_2 = 1$ for behavioral-only estimates and $\delta_3 = .4$ for neural-only estimates. Again, we used the sample mode of ABCDE to obtain posterior estimates, running the algorithm for 3,000 iterations with the first 1,500 iterations discarded. Migration and purification steps were performed in the same way as in joint estimation.

6.3. Prior Specification

To complete the specification in the Bayesian framework, we must specify priors for each of the model parameters. As we had no a priori beliefs about the model parameters, we chose the following uninformative priors for joint estimation, behavioral-only estimation and neural-only estimation:

$$\begin{aligned} c_2 &\sim U(0,1), \\ a_1 &\sim U(0,1), \\ \theta_1 &\sim U(0,1000), \\ \sigma_1 &\sim U(0,100), \text{ and} \\ A_{12} &\sim U(0,1), \end{aligned}$$

where $U(a, b)$ denotes a uniform distribution with lower bound a and upper bound b.

6.4. Results

Fig. 11 shows a comparison of the estimated posterior densities for parameters c_2, a_1, θ_1, σ_1 and A_{12} by the 36,000 posterior samples. The blue,

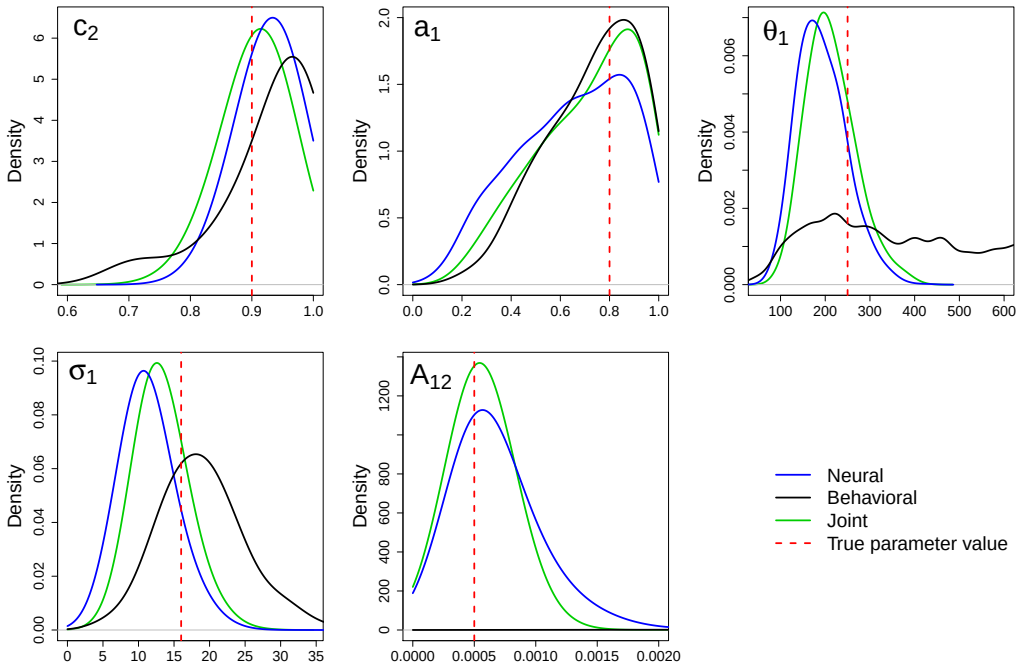

Figure 11: **Comparison of posterior estimates.** The estimated posterior density for each parameter informed by only neural data, only behavioral data, or jointly neural and behavioral data. Densities are smoothed using Gaussian kernels with widths .2, .5, 100, 10 and .0005 for parameter c_2, a_1, θ_1, σ_1 and A_{12} for illustration, respectively. The true parameter values are indicated by dashed red lines.

black and green lines indicate the posterior densities for each parameter informed by behavioral data-only, neural data-only, and jointly informed by behavioral and neural data, respectively. The red dashed lines indicate the true parameter values. Across all five parameters, the neural and joint estimates are near their true value. While estimates having only behavioral data are generally worse than the other modalities, the posteriors still contain the true parameter value (except for A_{12}). For parameter c_2, the neural density and joint density cover similar range but the joint density is closer to the true parameter value compared to the neural density. The behavioral density, on the contrary, has larger range of support for the posterior density and is more distant from the true parameter value compared with the other two. For parameter a_1, the neural density, behavioral density and joint density all include the true parameter value, whereas the behavioral density and joint density have higher peaks. For parameter θ_1, the joint density is close to the neural density, but the joint density is closer to the true parameter value. The behavioral density is more flat compared to the other two. For parameter σ_1, the joint density and neural density are close, but still the joint density is closer to the true parameter value. The behavioral density shows the best recovery performance for σ_1 among three, which is possible due to the critical importance on controlling signal-to-noise ratio of σ_1. The last parameter A_{12} differs from all the other four as it does not have a direct influence on behavioral data in the simulation process. Hence, the behavioral density is shown as a horizontal line. The neural and joint densities are close and both contain the true parameter value around their peaks. The joint density, however, has a slightly higher peak than the neural density and this trend can be due to the correlation between free parameters - correlations between free parameters make the estimation of one parameter able to inform the estimation of other parameters (Turner et al., 2019a). This comparison of posterior densities suggests the benefit by including both behavioral and neural data.

Fig. 12 shows the estimated posterior distributions informed by both behavioral and neural data for parameters c_2, a_1, θ_1, σ_1 and A_{12} by the 36,000 posterior samples. The panels on the diagonal show the marginal posterior distributions, where a dashed red vertical line indicates the true parameter value that was used to generate the observed data, and the dashed blue vertical line indicates the mean of the posterior estimates. All marginal posterior distributions deviate from their respective uniform priors, suggesting that the likelihood approximation is affecting the estimates. c_2 and a_1 are

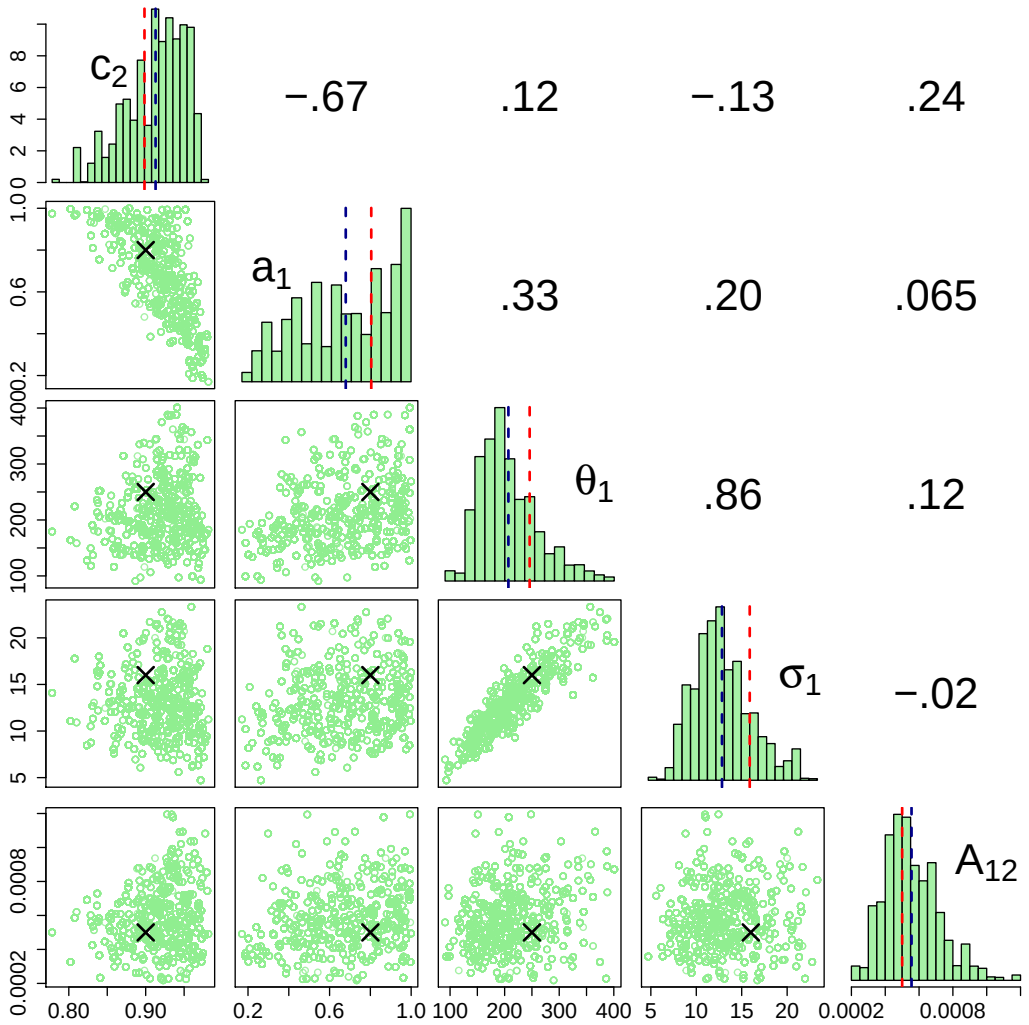

Figure 12: **Estimated joint posterior distributions.** The estimated marginal posterior distribution for each parameter is shown on the diagonal entries, whereas the estimated pairwise joint posterior distributions are shown in the bottom triangle. The upper triangle shows the correlation of the corresponding joint posterior distribution. In each panel containing a marginal estimate, a dashed red vertical line indicates the true parameter value used to generate the observed data, whereas a dashed blue vertical line indicates the posterior mean value. In each panel containing a joint estimate, a black "x" symbol denotes the location of the true parameter value.

both left-skewed, and this skewness is likely as their prior distributions are constrained to be less than 1. θ_1, σ_1 and A_{12} all have right-skewed posterior distributions. All five posterior means are close to the true parameter values. Each parameter estimate is well constrained and unimodal (except for some irregularities in a_1), suggesting that the model is securely identifiable.

The bottom left of the diagonal in Fig. 12 displays the pairwise joint posterior distributions between all five parameters, where the $x-$ and $y-$ axes can be inferred from the marginals. In each panel, the black "x" symbol indicates the true value of the parameter that generated the data. The top right of the diagonal displays the pairwise correlation coefficients. Combining the correlation plots and coefficients, we observe a strong negative correlation between c_2 and a_1 and a strong positive correlation between θ_1 and σ_1. These strong correlations are interpretable under model specifications. Recall that c_2 represents the within-region connection strength of R_3 and R_4 and a_1 represents connection strength from R_1 to R_3 and from R_2 to R_4. Therefore a reduction in a_1 should occur with an increase in c_2 so that both R_3 and R_4 can still accumulate the same amount of evidence. Regarding the positive correlation between θ_1 and σ_1, θ_1 is the threshold value that the absolute difference of neuronal activations that R_3 and R_4 accumulate toward and σ_1 controls the signal-to-noise ratio of the system. When σ_1 increases, the neuronal activation variations increase. The threshold value also needs to be higher, otherwise the neuronal activation variations could easily reach the threshold by random.

As a short conclusion, the parameter recovery study has suggested at least partial identifiability of the MDS model structure by estimating some important model parameters. The likelihood free algorithm KABC contributes to the parameter recovery of the MDS model. The posterior distribution of those estimated parameters captures the true parameter values and the posterior means are close to the true values. Hence the recovered parameters are veridical.

7. Discussion

In the present article, we have proposed and investigated a new framework for simultaneously modeling neural and behavioral data. Theoretically, it differs from the previous simultaneous modeling attempts in that both neural and behavioral data are linked by the same generative process, rather than linking them through an agnostic, parametric transformation. This theoret-

ical distinction produces a statistical distinction in that integrative models are more closely connected to the data. Whereas covariance approaches assume conditional independence between model parameters and data (i.e., by using different "submodels"), the integrative approach is directly committed to both streams of data: changes in a single parameter will affect the model predictions for both neural and behavioral data. The outcome of this direct connection is that it enables more precise model parameter estimates, as was shown when comparing the integrative model to models that only considered behavior or neural data. Although we have shown the utility of the MDS framework in two simulation studies and a parameter recovery study, there are a number of extensions and possibilities that we did not explore in the present article. In the following sections, we will discuss a few of these open questions and relationships as well as relating our work to previous efforts.

7.1. Comparison with DCM/MDS

Our extended MDS departs from other DCM/MDS models (Daunizeau et al., 2009; Friston et al., 2003; Marreiros et al., 2008; Ryali et al., 2011; Stephan et al., 2008, 2010) in several important ways. First, it incorporates the standard sequential sampling assumptions prevalent in extant models of evidence accumulation models to generate predictions for behavioral data. The self-connection parameters in the accumulation nodes (i.e. $C[3,3], C[4,4]$) are constrained to be less than and close to 1, which is analogous to having a "leakage" term often used in accumulator models (McClelland, 1993; Smith, 1995; Usher and McClelland, 2001). The two threshold parameters θ_1 and θ_2 are analogous to the threshold term commonly used in evidence accumulation models (for a review, see Ratcliff and Smith, 2004). However, deciding which subset of ROI(s) represent the accumulation of evidence is a nontrivial problem. In perceptual decision making, the LIP and FEF regions are well-established areas that may reflect the accumulation mechanism. For extensions of the model presented here to other cognitive processes, different configurations of the accumulation process may need to be considered. For example, building in a separate valuation process to represent the subjective strength of hedonic stimuli may need to operate prior to, or integrated within, the accumulation process described here (e.g., Turner et al., 2018). Compared to the behavior DCM approach (Rigoux and Daunizeau, 2015) where behavioral responses are predicted by a sigmoid mapping function of the latent neuronal activations, our approach allows continuous response times rather than only binary response choices. More importantly, our approach

provides mechanistic explanations for cognitive processes with interpretable model parameters (e.g. leakage, threshold, non-decision time).

Second, we have relaxed the connectivity parameters from being fixed throughout the time course to being temporally variant. In particular, we allow the endogenous connectivity matrix $C(t)$ to change from C_1 to C_2 after threshold-crossing time t_0 and back to C_1 after movement initiation time t_1 in each experimental trial. This change adds a complex nonlinearity to the MDS model, making it analytically intractable. For this reason, we recommend using fixed connectivity matrices as an initial exploratory step, and only allowing the connectivity matrices to change if there are explicit justifications for doing so. Such a tendency toward parsimony is productive in that it reduces the number of parameters that need to be estimated, provides strong constraints on the model, and helps to reduce any potential overfitting tendency in the model. Furthermore, the time points t_0 and t_1 for these changes to take place are determined by the interplay between the state equation Eq. 1 and the accumulation process in R_3 and R_4. Therefore, the accumulation process has a direct effect on the endogenous connectivity matrix and hence on the neural prediction. After t_1 when the response is initiated from R_6, $C(t)$ is changed from C_2 to C_1 to reflect a "resting stage" where neurons are prepared for the next trial. Except for the connectivity matrix, the noise term of neuronal activations in R_1 and R_2 drops from σ_1 to σ_2 (i.e., $\sigma_1 > \sigma_2$) after t_1, which also reflects the resting stage after the response is initiated from R_6. The inputs U_L and U_R are set to be zero after t_1 to indicate the termination of processing visual stimuli at this time point. By introducing such changes, we intend to consider underlying cognitive processes, behavioral responses and neural activities as a whole, rather than map one as a transformation of another. Hence, our framework can be thought of as an integrative approach to modeling behavioral and neural data simultaneously (Turner et al., 2017b, 2019a).

As an early attempt to explicitly modeling behavioral data (choice and response time) as well as fMRI BOLD signal, we did not include modulatory terms or nonlinear terms in the state equation Eq. 1 compared with previous efforts (Friston et al., 2003; Marreiros et al., 2008; Ryali et al., 2011; Stephan et al., 2008). On the other hand, the model is enriched by incorporating parameters representing an accumulation to bound process, and so there is not presently a clear conclusion about the complexity of our model relative to others. The current model specification might be thought of as a new, mechanistic version of DCM/MDS, where it is capable of explaining

behavioral data as an extra benefit.

Another difference between our extended MDS framework and DCM is that DCM facilitates model comparisons based on model evidence, so that different hypotheses about the connections among brain regions and how external input affects their interactions can be tested. Although we have not explicitly provided guidelines for how MDS could enable model comparison, we expect that analogous comparisons are easily made. For example, to avoid issues like model misspecification, one can directly compare a model that is intentionally misspecified to one that is not expected to be. Once each of these models are fit to data, one can simply compare the quality of those model fits. Models that are misspecified are expected to mismatch the pattern of data by design, so if they do not, then one can conclude that the data are not sufficiently able to identify models that are misspecified from those that are not. A more complex alternative is to test new mechanisms in the model by making a set of models with different mechanistic assumptions. For example, one could compare the baseline MDS model presented here to another model that includes a lateral inhibition mechanisms between R_1 and R_2, or between R_3 and R_4 (e.g., Ashby et al., 2007; Usher and McClelland, 2001). The addition of the inhibition mechanism would need to be justified by fitting to the data better than a model without inhibition, and the assessment of justification can easily be made by existing model performance metrics that balance fit to data with penalty terms for model complexity. It would also be possible to compare models with different configurations of $C(t)$ to guide decisions about how flexible the connectivity matrix should be in the time course of a cognitive process. In summary, although we didn't compare many different MDS models, we recommend that MDS can be used as a way to instantiate several different hypotheses within a computational model, where the models' fit to data, balanced for complexity, can be used to provide support for specific hypotheses about how the brain produces behavior.

7.2. ROI definition and identification

We defined six different ROIs in the perceptual decision making MDS model throughout the article. Here we discuss the possibility of identifying those ROIs from real fMRI data and some potential issues with defining and identifying the set of ROIs.

Theoretically, R_1 and R_2 can be identified by using MVPA methods and the tuning curve property of neurons within visual cortex (especially MT and MST; Kamitani and Tong, 2005, 2006; Serences and Boynton, 2007a,b).

However, in practice, the ability of MVPA for this purpose remains controversial. For example, MVPA classifiers may not find all the voxels that are relevant to represent the feature values, as they tend to overweight the importance of voxels that provide discriminative information and underweight voxels that are common to both (Norman et al., 2006). We have constrained R_1 and R_2 to only encode the properties of the external stimulus through U_L and U_R, respectively, but this constraint could be relaxed to construct more realistic models.

The nodes R_3 and R_4 are assumed to stand for separate voxels inside FEF and LIP. We treated FEF and LIP as a single ROI due to the similar functional roles of LIP and FEF in the perceptual decision making task. This assumption might lead to some issues, as it implies equal self-connectivity within FEF and LIP, which has not been supported by empirical results. Future investigations should consider MDS models with separate nodes representing similar functional roles. Also, we assumed that there are separate voxels responsible for integrating leftward and rightward motion information within FEF and LIP, but this assumption is not widely accepted.

Furthermore, the typical spatial resolution of fMRI may not be able to locate the output nuclei of basal ganglia (i.e. R_5). Although our framework assumes that neurons inside of an ROI carry homogeneous functions and share the same neuronal activations – a common assumption in cognitive neuroscience – the functional homogeneity of voxels inside an ROI has been shown to vary across ROIs and change in time (Korhonen et al., 2017; Ryyppö et al., 2018). Hence, inhomogeneity of voxels within an ROI will create a significant challenge to the static node definition used here.

Following the direction of information transformation in Fig. 3, there are four layers that contain multiple nodes, from visual cortex to pre SMA downstream. Thus the MDS framework can be viewed as a variant of neural network models, and it can be generalized to look more similar to neural network models by adding more units to each layer. This direction of generalization is reasonable, as the overall average activation of each ROI may be insufficient to represent the neural information contained in the ROIs, according to pattern-based information representation. Ideally, we can further parcelize ROIs into multiple nodes and use connectivity matrices for nodes between two layers, instead of scalar weights. The generalization requires overcoming at least two major difficulties though. First, when each layer contains more than two nodes, it is much harder to find the corresponding neural voxels for each node in that layer, and so it increases the complexity to

generate neural predictions for each node. The second difficulty comes from the well-known identifiability challenge and overfitting issue in neural network models. Allowing more nodes in each layer and connectivity matrices inevitably hinders the possibility of the model being identifiable.

7.3. Methods for Parameter Estimation

In this article, we have also investigated parameter recovery. To fit the model to data, we combined Bayesian MCMC posterior sampling with a kernel-based likelihood approximation method, known as kernel-based ABC (Palestro et al., 2018b). The kernel-based ABC method gives an approximation of the likelihood by considering summary statistics of three time series quantities: one for behavioral choice, one for behavioral response time, and one for the set of neural activations in the model. Kernel-based ABC techniques have the downside of having "tolerance" parameters where predictions of the model are compared to the observed data by measuring the discrepancy between them through a localized regression technique (Beaumont, 2010). Hence, while posteriors can be perfectly recovered with the discrepancy of the residual term is zero, it is often impossible for the residuals to be perfectly zero. This implies that any posterior approximation will have some error (e.g., have some increased variance) relative to the true posterior.

An alternative to this approach is the probability density approximation (PDA; Miletić et al., 2017; Molloy et al., 2019; Turner and Sederberg, 2014; Turner et al., 2015a; Turner and Van Zandt, 2018) method. Essentially, the PDA method relies on numerous simulations of the model for a candidate set of parameters to approximate the likelihood function through a kernel density estimation procedure (KDE; Silverman, 1986). The PDA method assumes a nonparametric form of the likelihood function whereas the kernel-based ABC method is based on the normal approximation, and so PDA often will provide a more accurate approximation of the likelihood function. As a downside, the PDA method is usually time-consuming due to the high number of model simulations often necessary for improving the likelihood approximation.

Another alternative for model fitting within the Bayesian framework is to use Variational Bayes to compute the posterior distributions of model parameters (e.g., Ryali et al., 2011). The Variational Bayes approach is able to obtain a posterior distribution of latent states and model parameters. Instead of relying on Monte Carlo property in MCMC sampling. To do so, Variational Bayes assumes a parametric form of the posterior distribution

and uses an iterative procedure to estimate the posteriors by minimizing the distance between the posterior distribution and the evolving parametric form (Galdo et al., 2019). Variational Bayes has been successful in fitting many other MDS/DCM models (Daunizeau et al., 2014; David et al., 2006; Friston et al., 2003; Marreiros et al., 2008; Ryali et al., 2011, 2016). However, Variational Bayes usually requires a known likelihood function, whereas the likelihood in the current model is intractable. It is presently unclear how Variational Bayesian methods will perform when optimizing over highly stochastic gradients.

7.4. Limitations and future directions

7.4.1. Choice of features

Our example random dot motion task is based on a widely studied low-level feature: motion direction. Numerous monkey and human studies have shown the existence of separate neurons or voxels sensitive to each motion-direction, and thus we believe it is entirely possible to identify R_1 and R_2 from various fMRI voxels based on MVPA techniques. Other than motion direction, many low-level physical features have investigated the encoding properties (e.g. tuning curves) of voxels, such as line orientation, color, and spatial location. However, we have not accumulated sufficient knowledge about the properties of neural encoding for many higher-level or abstract features. For example, it is not clear if voxels can specifically code for smaller-sooner versus larger-later options in intertemporal choice, or for preferences among food options.

Before knowing how higher-level feature values are encoded in individual voxels, it is reasonable to remain conservative and apply this approach to tasks that are based on low-level physical features. Note that limiting investigations to low-level features does not limit the scope of applying this framework to low-level cognitive problems, because many higher-level cognitive problems (e.g., memory, categorization) can be investigated with stimuli using low-level features.

7.4.2. Number of parameters in the recovery study

In the parameter recovery study, we chose to recover only 5 of the 18 parameters listed in Table 2. Therefore, we cannot guarantee that the whole model structure is identifiable. It is expected that many more iterations would be required to estimate the full model, and even then, we consider it unlikely that all model parameters will be well recovered without significant

amounts of data. The number of required iterations for convergence also depends on which parameters are chosen to estimate, and the starting values of chains for the estimation. More in-depth exploration is required to find the effects of different numbers and different configurations of free parameters on the computational costs of parameter estimation, as well as the particular set of experimental and data constraints that will ensure parameter identifiability.

7.4.3. Simulated fMRI design

For the assumed rapid event-related fMRI design in the simulation, we used a constant inter-trial-interval (ITI) to reduce the complexity of model simulation. However, a jittered ITI is more commonly adopted in real rapid event-related designs as a way to minimize confounds from a subject's habituation, as well as increasing the efficiency of estimating the hemodynamic impulse response based on the periodic overlap among stimulus-related hemodynamic functions (Birn et al., 2002; Liu et al., 2001). The model has to be further refined so that the simulated fMRI data can be more comparable with real fMRI data.

7.4.4. Modality of neural data modeling

We have mapped the neuronal activations to fMRI time series data via a linear convolution with a canonical HRF, but under the temporal resolution of fMRI, our framework may be better situated to model EEG time series data. We simulated the time series of neuronal activations on the millisecond level, but when mapping the neural activations to the fMRI BOLD signal, we had to downsample the simulated fMRI BOLD signal by a factor of 1,000 to mimic the real sampling resolution of typical fMRI signals. In so doing, we have discarded significant information about the temporal dynamics of our model. On the other hand, EEG data can easily achieve a temporal resolution of 1 millisecond. In fact, DCM has been extended to generate EEG/MEG data by use of a neural mass model to spatially map the unobserved neuronal activations to the EEG/MEG evoked responses (David et al., 2006; Kiebel et al., 2008). Another opportunity is to use anatomical sources from fMRI to constrain source localization methods for EEG data. Such efforts could exploit the temporal resolution of EEG and the spatial resolution of fMRI to form a more complete picture of brain dynamics (Turner et al., 2016). Although we are currently working on including EEG measurements in the

generative model to take advantage of temporal information, such efforts were beyond the initial scope of this article.

8. Conclusion

Our results suggest that the extended MDS framework may prove useful for future efforts in developing fully integrated models of brain and behavior. We have shown that integrated models can be used to produce patterns of neural and behavioral data that resemble experimental results. We have also shown that we can recover the model parameters when fit to simulated data, where the true values of the model parameters are known. Together, these results suggest that MDS may be productive in inferring causal links that explain how behavior may emerge from the brain through mental operations.

9. References

Albright, T. D., 1984. Direction and orientation selectivity of neurons in visual area mt of the macaque. Journal of neurophysiology 52 (6), 1106–1130.

Andersen, R. A., Brotchie, P. R., Mazzoni, P., 1992. Evidence for the lateral intraparietal area as the parietal eye field. Current opinion in neurobiology 2 (6), 840–846.

Anderson, J. R., 2007. How can the human mind occur in the physical universe? Oxford University Press, New York, NY.

Anderson, J. R., Byrne, D., Fincham, J. M., Gunn, P., 2008. Role of prefrontal and parietal cortices in associative learning. Cerebral Cortex 18, 904–914.

Ashby, F. G., Ennis, J. M., Spiering, B. J., 2007. A neurobiological theory of automaticity in perceptual categorization. Psychological review 114 (3), 632.

Ball, K., Sekuler, R., 1982. A specific and enduring improvement in visual motion discrimination. Science 218 (4573), 697–698.

Bamber, D., Van Santen, J. P., 2000. How to assess a model's testability and identifiability. Journal of Mathematical Psychology 44 (1), 20–40.

Beaumont, M. A., 2010. Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics 41, 379–406.

Birn, R. M., Cox, R. W., Bandettini, P. A., 2002. Detection versus estimation in event-related fmri: choosing the optimal stimulus timing. Neuroimage 15 (1), 252–264.

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., Nieuwenhuis, S., 2010. The neural basis of the speed-accuracy tradeoff. Trends in Neuroscience 33, 10–16.

Borst, J. P., Anderson, J. R., 2013. Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the

fronto-parietal network. Proceedings of the National Academy of Sciences of the United States 110, 1628–1633.

Borst, J. P., Taatgen, N. A., Stocco, A., Van Rijn, H., 2010a. The neural correlates of problem states: Testing fMRI predictions of a computational model of multitasking. PLoS ONE 5, e12966.

Borst, J. P., Taatgen, N. A., Van Rijn, H., 2010b. The problem state: A cognitive bottleneck in multitasking. Journal of Experimental Psychology: Learning, Memory, & Cognition 36, 363–382.

Boucher, L., Palmeri, T. J., Logan, G. D., Schall, J. D., 2007. Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychological review 114 (2), 376.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., Movshon, J. A., 1996. A relationship between behavioral choice and the visual responses of neurons in macaque mt. Visual neuroscience 13 (1), 87–100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., Movshon, J. A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. The Journal of Neuroscience 12 (12), 4745–4765.

Brown, J. W., Hanes, D. P., Schall, J. D., Stuphorn, V., 2008. Relation of frontal eye field activity to saccade initiation during a countermanding task. Experimental Brain Research 190 (2), 135.

Brown, S., Heathcote, A., 2005. A ballistic model of choice response time. Psychological Review 112, 117–128.

Brown, S., Heathcote, A., 2008. The simplest complete model of choice reaction time: Linear ballistic accumulation. Cognitive Psychology 57, 153–178.

Buxton, R. B., Wong, E. C., Frank, L. R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magnetic resonance in medicine 39 (6), 855–864.

Carpenter, R., 1999. Visual selection: Neurons that make up their minds. Current Biology 9 (16), R595–R598.

Carpenter, R., Reddi, B., Anderson, A., 2009. A simple two-stage model predicts response time distributions. The Journal of physiology 587 (16), 4051–4062.

Carpenter, R. H., Williams, M., 1995. Neural computation of log likelihood in control of saccadic eye movements. Nature 377 (6544), 59.

Cassey, P. J., Gaut, G., Steyvers, M., Brown, S. D., 2016. A generative joint model for spike trains and saccades during perceptual decision-making. Psychonomic bulletin & review 23 (6), 1757–1778.

Celebrini, S., Newsome, W. T., 1995. Microstimulation of extrastriate area mst influences performance on a direction discrimination task. Journal of Neurophysiology 73 (2), 437–448.

Churchland, A. K., Kiani, R., Shadlen, M. N., 2008. Decision-making with multiple alternatives. Nature neuroscience 11 (6), 693.

Colby, C. L., Goldberg, M. E., 1999. Space and attention in parietal cortex. Annual review of neuroscience 22 (1), 319–349.

Croner, L. J., Albright, T. D., 1999. Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. Journal of Neuroscience 19 (10), 3935–3951.

Daunizeau, J., Adam, V., Rigoux, L., 2014. Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS computational biology 10 (1), e1003441.

Daunizeau, J., Friston, K. J., Kiebel, S. J., 2009. Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models. Physica D: nonlinear phenomena 238 (21), 2089–2118.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., Friston, K. J., 2006. Dynamic causal modeling of evoked responses in eeg and meg. NeuroImage 30 (4), 1255–1272.

de Hollander, G., Forstmann, B. U., Brown, S. D., 2016. Different ways of linking behavioral and neural data via computational cognitive models. Cognitive Neuroscience and Neuroimaging 1, 101–109.

Ding, L., Gold, J. I., 2013. The basal ganglia's contributions to perceptual decision making. Neuron 79 (4), 640–649.

Dorris, M. C., Pare, M., Munoz, D. P., 1997. Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. Journal of Neuroscience 17 (21), 8566–8579.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., Turner, R., 2010. Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. Proceedings of the National Academy of Sciences 107 (36), 15916–15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., Wagenmakers, E.-J., 2008. Striatum and pre-SMA facilitate decision-making under time pressure. Proceedings of the National Academy of Sciences 105 (45), 17538–17542.

Forstmann, B. U., Wagenmakers, E.-J., 2015. An introduction to model-based cognitive neuroscience. Springer, New York.

Frigo, M., Johnson, S. G., 2005. The design and implementation of FFTW3. Proceedings of the IEEE 93 (2), 216–231, special issue on "Program Generation, Optimization, and Platform Adaptation".

Friston, K., 2009. Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS biology 7 (2), e1000033.

Friston, K., Harisson, L., Penny, W., 2003. Dynamic causal modeling. NeuroImage 19, 1273–1302.

Friston, K., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., Zeidman, P., 2017. Dynamic causal modelling revisited. NeuroImage.

Friston, K. J., Mechelli, A., Turner, R., Price, C. J., 2000. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. NeuroImage 12 (4), 466–477.

Galdo, M., Bahg, G., Turner, B. M., 2019. Variational bayesian methods for cognitive science, in press at Psychological Methods.

Georgiev, D., Rocchi, L., Tocco, P., Speekenbrink, M., Rothwell, J. C., Jahanshahi, M., 2016. Continuous theta burst stimulation over the dorsolateral prefrontal cortex and the pre-sma alter drift rate and response thresholds respectively during perceptual decision-making. Brain stimulation 9 (4), 601–608.

Gold, J. I., Shadlen, M. N., 2001. Neural computations that underlie decisions about sensory stimuli. Trends in cognitive sciences 5 (1), 10–16.

Gold, J. I., Shadlen, M. N., 2002. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36 (2), 299–308.

Gold, J. I., Shadlen, M. N., 2007. The neural basis of decision making. Annual Review of Neuroscience 30, 535–574.

Graybiel, A. M., 1995. Building action repertoires: memory and learning functions of the basal ganglia. Current opinion in neurobiology 5 (6), 733–741.

Hikosaka, O., Nakamura, K., Nakahara, H., 2006. Basal ganglia orient eyes to reward. Journal of neurophysiology 95 (2), 567–584.

Hikosaka, O., Takikawa, Y., Kawagoe, R., 2000a. Role of the basal ganglia in the control of purposive saccadic eye movements. Physiological reviews 80 (3), 953–978.

Hikosaka, O., Takikawa, Y., Kawagoe, R., 2000b. Role of the basal ganglia in the control of purposive saccadic eye movements. Physiological reviews 80 (3), 953–978.

Ho, T. C., Brown, S., Serences, J. T., 2009. Domain general mechanisms of perceptual decision making in human cortex. Journal of Neuroscience 29 (27), 8675–8687.

Holmes, W. R., 2015. A practical guide to the probability density approximation (pda) with improved implementation and error characterization. Journal of Mathematical Psychology 68, 13–24.

Houk, J. C., Davis, J. L., Beiser, D. G., 1995. Models of information processing in the basal ganglia. MIT press.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nature neuroscience 8 (5), 679.

Kamitani, Y., Tong, F., 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Current biology 16 (11), 1096–1102.

Kiebel, S. J., Garrido, M. I., Moran, R. J., Friston, K. J., 2008. Dynamic causal modelling for eeg and meg. Cognitive neurodynamics 2 (2), 121.

Kim, J. N., Shadlen, M. N., 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nature Neuroscience 2, 176–185.

Korhonen, O., Saarimäki, H., Glerean, E., Sams, M., Saramäki, J., 2017. Consistency of regions of interest as nodes of fmri functional brain networks. Network Neuroscience 1 (3), 254–274.

Kragel, J. E., Morton, N. W., Polyn, S. M., 2015. Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. Journal of Neuroscience 35 (7), 2914–2926.

Liu, T. T., Frank, L. R., Wong, E. C., Buxton, R. B., 2001. Detection power, estimation efficiency, and predictability in event-related fmri. Neuroimage 13 (4), 759–773.

Lo, C.-C., Wang, X.-J., 2006. Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nature neuroscience 9 (7), 956.

Mandeville, J. B., Marota, J. J., Ayata, C., Zaharchuk, G., Moskowitz, M. A., Rosen, B. R., Weisskoff, R. M., 1999. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. Journal of Cerebral Blood Flow & Metabolism 19 (6), 679–689.

Mansfield, E. L., Karayanidis, F., Jamadar, S., Heathcote, A., Forstmann, B. U., Oct 2011. Adjustments of response threshold during task switching: a model-based functional magnetic resonance imaging study. J Neurosci 31 (41), 14688–92.

Marreiros, A. C., Kiebel, S. J., Friston, K. J., 2008. Dynamic causal modelling for fmri: a two-state model. Neuroimage 39 (1), 269–278.

Maunsell, J. H., Van Essen, D. C., 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed, and orientation. Journal of neurophysiology 49 (5), 1127–1147.

McClelland, J. L., 1993. Toward a theory of information processing in graded, random, interactive networks. In: Meyer, D. E., Kornblum, S. (Eds.), Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience. Cambridge, MA: MIT Press, pp. 655–688.

Miletić, S., Turner, B. M., Forstmann, B. U., van Maanen, L., 2017. Parameter recovery for the leaky competing accumulator model. Journal of Mathematical Psychology 76, 25–50.

Molloy, M. F., Galdo, M., Bahg, G., Liu, Q., Turner, B. M., 2019. What?s in a response time?: On the importance of response time measures in constraining models of context effects. Decision 6 (2), 171.

Niwa, M., Ditterich, J., 2008. Perceptual decisions between multiple directions of visual motion. Journal of Neuroscience 28 (17), 4435–4445.

Norman, K. A., Polyn, S. M., Detre, G. J., Haxby, J. V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends in cognitive sciences 10 (9), 424–430.

O'Reilly, R. C., 2006. Biologically based computational models of cortical cognition. Science 314, 91–94.

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., Turner, B. M., 2018a. A tutorial on joint models of neural and behavioral measures of cognition. Journal of Mathematical Psychology 84, 20–48.

Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., Turner, B. M., 2018b. Likelihood-free methods for cognitive science. Springer.

Penny, W., Ghahramani, Z., Friston, K., 2005. Bilinear dynamical systems. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1457), 983–993.

Pirrone, A., Stafford, T., Marshall, J. A., 2014. When natural selection should optimize speed-accuracy trade-offs. Frontiers in neuroscience 8, 73.

Polyn, S. M., Natu, V. S., Cohen, J. D., Norman, K. A., 2005. Category-specific cortical activity precedes retrieval during memory search. Science 310 (5756), 1963–1966.

Purcell, B., Heitz, R., Cohen, J., Schall, J., Logan, G., Palmeri, T., 2010. Neurally-constrained modeling of perceptual decision making. Psychological Review 117, 1113–1143.

Ratcliff, R., Cherian, A., Segraves, M., 2003. A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. Journal of Neurophysiology 90, 1392–1407.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Smith, P. L., Segraves, M. A., 2007. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. Journal of Neurophysiology 97, 1756–1774.

Ratcliff, R., Rouder, J. N., 1998. Modeling response times for two-choice decisions. Psychological Science 9, 347–356.

Ratcliff, R., Smith, P. L., 2004. A comparison of sequential sampling models for two-choice reaction time. Psychological Review 111, 333–367.

Ratcliff, R., Voskuilen, C., Teodorescu, A., 2018. Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. Cognitive psychology 103, 1–22.

Redgrave, P., Prescott, T. J., Gurney, K. N., 1999. The basal ganglia: A vertebrate solution to the selection problem? Neuroscience 89 (4), 1009–1023.

Rigoux, L., Daunizeau, J., 2015. Dynamic causal modelling of brain–behaviour relationships. Neuroimage 117, 202–221.

Roe, R. M., Busemeyer, J. R., Townsend, J. T., 2001. Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review 108, 370–392.

Roitman, J., Shadlen, M., 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. Journal of Neuroscience 22 (21), 9475–9489.

Ryali, S., Chen, T., Supekar, K., Tu, T., Kochalka, J., Cai, W., Menon, V., 2016. Multivariate dynamical systems-based estimation of causal brain interactions in fmri: Group-level validation using benchmark data, neurophysiological models and human connectome project data. Journal of neuroscience methods 268, 142–153.

Ryali, S., Supekar, K., Chen, T., Menon, V., 2011. Multivariate dynamical systems models for estimating causal interactions in fmri. Neuroimage 54 (2), 807–823.

Ryyppö, E., Glerean, E., Brattico, E., Saramäki, J., Korhonen, O., 2018. Regions of interest as nodes of dynamic functional brain networks. Network Neuroscience 2 (4), 513–535.

Salzman, C. D., Newsome, W. T., 1994. Neural mechanisms for forming a perceptual decision. Science 264 (5156), 231–237.

Schall, J. D., 2003. Neural correlates of decision processes: neural and mental chronometry. Current Opinion in Neurobiology 12, 182–186.

Schall, J. D., Morel, A., King, D. J., Bullier, J., 1995. Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. Journal of Neuroscience 15 (6), 4464–4487.

Serences, J. T., Boynton, G. M., 2007a. Feature-based attentional modulations in the absence of direct visual stimulation. Neuron 55 (2), 301–312.

Serences, J. T., Boynton, G. M., 2007b. The representation of behavioral choice for motion in human visual cortex. Journal of Neuroscience 27 (47), 12893–12899.

Shadlen, M. N., Britten, K. H., Newsome, W. T., Movshon, J. A., 1996. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. Journal of Neuroscience 16 (4), 1486–1510.

Shadlen, M. N., Newsome, W. T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of Neurophysiology 86, 1916–1936.

Silverman, B. W., 1986. Density estimation for statistics and data analysis. London: Chapman & Hall.

Simoncelli, E. P., Heeger, D. J., 1998. A model of neuronal responses in visual area mt. Vision research 38 (5), 743–761.

Smith, J. F., Pillai, A., Chen, K., Horwitz, B., 2010. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. Neuroimage 52 (3), 1027–1040.

Smith, P. L., 1995. Psychophysically principled models of visual simple reaction time. Psychological review 102 (3), 567–593.

Smith, P. L., Vickers, D., 1988. The accumulator model of two-choice discrimination. Journal of Mathematical Psychology 32, 135–168.

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., Friston, K. J., 2008. Nonlinear dynamic causal models for fmri. Neuroimage 42 (2), 649–662.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., Friston, K. J., 2010. Ten simple rules for dynamic causal modeling. Neuroimage 49 (4), 3099–3109.

Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., Friston, K. J., 2007. Comparing hemodynamic models with dcm. Neuroimage 38 (3), 387–401.

Stewart, T. C., Choo, X., Eliasmith, C., 2010. Symbolic reasoning in spiking neurons: A model of the cortex/basal ganglia/thalamus loop. In: Catrambone, R., Ohlsson, S. (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin, TX, pp. 1100–1105.

Teller, D. Y., 1984. Linking propositions. Vision Research 24, 1233–1246.

Teodorescu, A. R., Moran, R., Usher, M., 2016. Absolutely relative or relatively absolute: violations of value invariance in human decision making. Psychonomic bulletin & review 23 (1), 22–38.

Teodorescu, A. R., Usher, M., 2013. Disentangling decision models – from independence to competition. Psychological Review 120, 1–38.

ter Braak, C. J. F., 2006. A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing 16, 239–249.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M. P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6, 187–202.

Turner, B., Wang, T., Merkle, E., 2017a. Factor analysis linking functions for simultaneously modeling neural and behavioral data. NeuroImage 153, 28–48.

Turner, B. M., 2019. Toward a common representational framework for adaptation, in press at Psychological Review.

Turner, B. M., Dennis, S., Van Zandt, T., 2013a. Bayesian analysis of memory models. Psychological Review 120, 667–678.

Turner, B. M., Forstmann, B. U., Love, B. U., Palmeri, T. J., Van Maanen, L., 2017b. Approaches to analysis in model-based cognitive neuroscience. Journal of Mathematical Psychology 76, 65–79.

Turner, B. M., Forstmann, B. U., Steyvers, M., et al., 2019a. Joint models of neural and behavioral data. Springer.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., Steyvers, M., 2013b. A bayesian framework for simultaneously modeling neural and behavioral data. NeuroImage 72, 193–206.

Turner, B. M., Palestro, J. J., Miletić, S., Forstmann, B. U., 2019b. Advances in techniques for imposing reciprocity in brain-behavior relations. Neuroscience & Biobehavioral Reviews 102, 327–336.

Turner, B. M., Rodriguez, C. A., Liu, Q., Molloy, M. F., Hoogendijk, M., McClure, S. M., 2018. On the neural and mechanistic bases of self-control. Cerebral cortex 29 (2), 732–750.

Turner, B. M., Rodriguez, C. A., Norcia, T., Steyvers, M., McClure, S. M., 2016. Why more is better: A method for simultaneously modeling EEG, fMRI, and behavior. NeuroImage 128, 96–115.

Turner, B. M., Sederberg, P. B., 2012. Approximate Bayesian computation with Differential Evolution. Journal of Mathematical Psychology 56, 375–385.

Turner, B. M., Sederberg, P. B., 2014. A generalized, likelihood-free method for parameter estimation. Psychonomic Bulletin and Review 21, 227–250.

Turner, B. M., Sederberg, P. B., Brown, S., Steyvers, M., 2013c. A method for efficiently sampling from distributions with correlated dimensions. Psychological Methods 18, 368–384.

Turner, B. M., Sederberg, P. B., McClelland, J. L., 2015a. Bayesian analysis of simulation-based models, in Press.

Turner, B. M., Van Maanen, L., Forstmann, B. U., 2015b. Informing cognitive abstractions with neurophysiology: The neural drift diffusion model. Psychological Review 122, 312–336.

Turner, B. M., Van Zandt, T., 2014. Hierarchical approximate Bayesian computation. Psychometrika 79, 185–209.

Turner, B. M., Van Zandt, T., 2018. Approximating bayesian inference through model simulation. Trends in cognitive sciences.

Tversky, A., Simonson, I., 1993. Context-dependent preferences. Management science 39 (10), 1179–1189.

Usher, M., McClelland, J. L., 2001. The time course of perceptual choice: The leaky competing accumulator model. Psychological Review 108, 550–592.

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., 2011. Neural correlates of trial-to-trial fluctuations in response caution. Journal of Neuroscience 31, 17488–17495.

van Ravenzwaaij, D., Provost, A., Brown, S. D., 2017. A confirmatory approach for integrating neural and behavioral data into a single model. Journal of Mathematical Psychology 76, 131–141.

Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., Tootell, R. B., Orban, G. A., 2001. Visual motion processing investigated using contrast agent-enhanced fmri in awake behaving monkeys. Neuron 32 (4), 565–577.

Wagenmakers, E.-J., Farrell, S., Ratcliff, R., 2004. Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. Psychonomic Bulletin and Review 11, 579–615.

Wickens, J., 1997. Basal ganglia: structure and computations. Network: Computation in Neural Systems 8 (4), R77–R109.

Zeki, S. M., 1974. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. The Journal of Physiology 236 (3), 549–573.

Algorithm 1 Pseudocode for MDS model

$I \leftarrow$ Number of trials
$T \leftarrow$ Number of time points in a trial
Specify $D, C_1, C_2, \sigma_1, \sigma_2, \theta_1, \theta_2, \tau$
for $1 \leq i \leq I$ **do**
    Specify U; Initialize $S \leftarrow \mathbf{0}, t_0 \leftarrow T, t_1 \leftarrow T$, choice $\leftarrow$ null
    if $i > 1$ **then**
        Use the $S(T)$ of trial $i-1$ as $S(1)$ of current trial
    end if
    for $2 \leq t \leq T$ **do**
        $Q(t) \leftarrow \text{diag}([\sigma_1, \sigma_1, \sigma_1, \sigma_1, \sigma_1, \sigma_1])$
        $\omega(t) \sim N(0, Q(t))$
        $S(t) \leftarrow C_1 S(t-1) + DU(t) + \omega(t)$
        if $\mid S_3(t) - S_4(t) \mid > \theta_1$ **then**
            Break current loop;
            $t_0 \leftarrow t$;
            winner $\leftarrow$ left if $S_3(t) > S_4(t)$, right if $S_4(t) > S_3(t)$
        end if
    end for
    if $t_0 < T$ **then**
        for $t_0 + 1 \leq t \leq T$ **do**
            $Q(t) \leftarrow \text{diag}([\sigma_1, \sigma_1, \sigma_1, \sigma_1, \sigma_1, \sigma_1])$
            $\omega(t) \sim N(0, Q(t))$
            $S(t) \leftarrow C_2 S(t-1) + DU(t) + \omega(t)$
            if $\sum_{n=1}^{t} S_6(n) > \theta_2$ **then**
                Break current loop;
                $t_1 \leftarrow t$;
                choice $\leftarrow$ winner
            end if
        end for
    end if
    if $t_1 < T$ **then**
        for $t_1 + 1 \leq t \leq T$ **do**
            $Q(t) \leftarrow \text{diag}([\sigma_2, \sigma_2, \sigma_1, \sigma_1, \sigma_1, \sigma_1])$
            $\omega(t) \sim N(0, Q(t))$
            $S(t) \leftarrow C_1 S(t-1) + \omega(t)$
        end for
    end if
    return $S, RT \leftarrow t_1 + \tau$, choice
end for
Concatenate S across trials
Perform convolution with HRF for each ROI
return fMRI BOLD activity