Psychological Review
2015, Vol. 122, No. 2, 312–336

0033-295X/15/$12.00 http://dx.doi.org/10.1037/a0038894

Informing Cognitive Abstractions Through Neuroimaging: The Neural Drift Diffusion Model

Brandon M. Turner
The Ohio State University

Leendert van Maanen and Birte U. Forstmann
University of Amsterdam

Trial-to-trial fluctuations in an observer's state of mind have a direct influence on their behavior. However, characterizing an observer's state of mind is difficult to do with behavioral data alone, particularly on a single-trial basis. In this article, we extend a recently developed hierarchical Bayesian framework for integrating neurophysiological information into cognitive models. In so doing, we develop a novel extension of the well-studied drift diffusion model (DDM) that uses single-trial brain activity patterns to inform the behavioral model parameters. We first show through simulation how the model outperforms the traditional DDM in a prediction task with sparse data. We then fit the model to experimental data consisting of a speed-accuracy manipulation on a random dot motion task. We use our cognitive modeling approach to show how prestimulus brain activity can be used to simultaneously predict response accuracy and response time. We use our model to provide an explanation for how activity in a brain region affects the dynamics of the underlying decision process through mechanisms assumed by the model. Finally, we show that our model performs better than the traditional DDM through a cross-validation test. By combining accuracy, response time, and the blood oxygen level–dependent response into a unified model, the link between cognitive abstraction and neuroimaging can be better understood.

Keywords: neural drift diffusion model, cognitive modeling, default mode network, evidence accumulation, neural correlates

Supplemental materials: http://dx.doi.org/10.1037/a0038894.supp

As psychologists, our ultimate goal is to fully understand how the mind produces behavior. However, the path to achieving this goal is riddled with obstacles that make our endeavor difficult, if not impossible. The challenge lies in the logistics of studying a highly flexible and dynamic system that is constantly evolving as a consequence of the task environment (cf. Criss, Malmberg, & Shiffrin, 2011; Logan, 1988; Turner, Van Zandt, & Brown, 2011). To make matters worse, the experimental data we use to understand this process may or may not even be cognitively relevant. For example, data obtained from a distracted or fatiguing subject may be inconsistent with the assumptions made in our particular cognitive model. Some would argue that these data completely invalidate our model, whereas others would simply treat these data as contaminants, effectively striking them from the analysis.

Given the ever-changing nature of the mind, perhaps the most comprehensive account of cognition would strive for trial-to-trial explanations of the mind's internal representation and how this representation might be used to generate behavior. Here we focus on how the dynamics of the mind's internal representations affect the decision-making behavior. There are many ways to incorporate trial-to-trial effects into cognitive models of decision making, including designating separate parameters for each trial (e.g., DeCarlo, 2011; Pratte & Rouder, 2011; Vandekerckhove, Tuerlinckx, & Lee, 2008, 2011; van Maanen et al., 2011), defining a statistical dependence on the basis of response choice or feedback (e.g., Craigmile, Peruggia, & Zandt, 2010; Kac, 1962, 1969; Peruggia, Van Zandt, & Chen, 2002; Treisman & Williams, 1984), or explicitly specifying how the task environment (e.g., the stimuli or an observer's responses) shapes an observer's representation over time (e.g., Criss et al., 2011; Howard & Kahana, 2002; Logan, 1988; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008; Turner et al., 2011; Vickers & Lee, 1998, 2000). Although all of these approaches have provided—in one way or another—a greater understanding of the dynamics of decision making, they are only designed to account for behavioral data. As a consequence, the insight they provide about an observer's state of mind on a given trial is limited to the abstractions assumed by the model. Furthermore, they can only provide

Brandon M. Turner, Psychology Department, The Ohio State University; Leendert van Maanen, Psychology Department, University of Amsterdam; Birte U. Forstmann, Amsterdam Brain and Cognition Department, University of Amsterdam.

This work was funded by National Institutes of Health award number F32GM103288 (Brandon M. Turner) and by a Vidi grant by the Dutch Organization for Scientific Research (NWO; Birte U. Forstmann), as well as a starter grant from the European Research Council (ERC; Birte U. Forstmann). Portions of this work were presented at the 12th Annual Summer Interdisciplinary Conference, Cortina d′Ampezzo, Italy. The authors thank Tom Eichele and Max Keuken for help in performing the fMRI data analysis and Andrew Heathcote, James McClelland, and Eric-Jan Wagenmakers for helpful comments that improved an earlier version of the manuscript. Data are available upon request.

Correspondence concerning this article should be addressed to Brandon M. Turner, The Ohio State University, 1827 Neil Avenue, Lazenby Hall, Room 200C, Columbus, OH 43210. E-mail: turner.826@gmail.com

predictions about behavioral performance on the basis of past behavior. That is, for a particular trial, these models are incapable of incorporating the observer's state of mind into predictions about subsequent behavioral outcomes.

We now have tools to examine an observer's state of mind at the neurophysiological level, through techniques such as functional MRI (fMRI) or electroencephalography (EEG). The importance of these measures in characterizing an observer's state of mind has been demonstrated by many authors (cf. Forstmann & Wagenmakers, 2014; Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011). In this article, we develop a new statistical approach for augmenting cognitive models with neurophysiological measures. Our approach extends the joint modeling framework (Turner, Forstmann, et al., 2013) to establish the first model of perceptual decision making that accounts for both neural and behavioral data at the single-trial level. The model allows us to study how trial-to-trial fluctuations in the pattern of neural data lead to systematic fluctuations in behavioral response patterns. We begin by presenting the conceptual and technical details of the model. We then show how the inclusion of trial-to-trial measures of neural activity can greatly affect the accuracy of the model's predictions relative to a model that captures only behavioral data. We then apply our model to data from a perceptual decision-making experiment, which allows us to interpret neurophysiological patterns on the basis of the mechanisms assumed by our cognitive model.

Integrating Neural and Behavioral Measures

An unsettling amount of what we know about human cognition has evolved from two virtually exclusive groups of researchers. The first group, known as mathematical psychologists, relies on a system of mathematical and statistical mechanisms to describe the cognitive process assumed to be underlying a decision. In an attempt to achieve parsimony and psychological interpretability, mathematical models are inherently abstract and rely on the estimation of latent model parameters to guide the inference process. The second group, known as cognitive neuroscientists, generally relies on statistical models (e.g., the general linear model) to determine whether an experimental manipulation produces a significant change in activity in a particular brain region.[1] Because this type of analysis makes no connection to an explicit cognitive theory, a mechanistic understanding of brain function cannot be achieved.

Both approaches suffer from critical limitations as a direct result of their focus on data at one level of analysis (cf. Marr, 1982). For example, without a cognitive theory to guide the inferential process, neuroscientists are (a) unable to interpret their results from a mechanistic point of view, (b) unable to address many phenomena with only contrast analyses (see, e.g., Todd, Nystrom, & Cohen, 2013), and (c) unable to explain results from different paradigms under a common theoretical framework. On the other hand, the cognitive models developed by mathematical psychologists are not informed by physiology or brain function. Instead, these researchers posit the existence of abstract mechanisms that are understood through the estimation of the model's parameters. For example, traditional sequential sampling models assume that the presentation of a stimulus gives rise to a race between decision alternatives to obtain a "threshold" amount of evidence. The race involves sequentially sampling evidence for each alternative at a rate dictated by another parameter, called the "drift rate." These models each make different assumptions about the types of variability that are present either between or within trials, but ultimately it is the estimate of the model parameters that serves as a proxy for the underlying decision dynamics.

Given the unavoidable limitations of both approaches, recent cognitive modeling endeavors have aimed at supporting cognitive theories by mapping the mechanistic explanations provided by cognitive models to the neural signal present in the data. The motivation for these efforts is clear: Neural data provide physiological signatures of cognition that inform the development of formal cognitive models (de Lange, Jensen, & Dehaene, 2010; de Lange, van Gaal, Lamme, & Dehaene, 2011; O'Connell, Dockree, & Kelly, 2012), providing greater constraint on cognitive theory than behavioral data alone (Forstmann, Wagenmakers, et al., 2011). Despite the utility of neural data, they are not the cure-all (e.g., Anderson et al., 2008). Without a motivating theory for *why* particular brain regions become active, interpretations regarding the functional role of brain regions can be difficult to substantiate. We argue that to fully understand cognition, the relationship between cognitive neuroscience and cognitive modeling must be reciprocal (Forstmann, Wagenmakers, et al., 2011).

In light of the advantages of cognitive models, several authors have used cognitive models in conjunction with neural measures, an approach we refer to as "model-based cognitive neuroscience" (Forstmann, Wagenmakers, et al., 2011). With some exceptions (Anderson, Betts, Ferris, & Fincham, 2010; Anderson et al., 2008; Anderson, Fincham, Schneider, & Yang, 2012; Mack, Preston, & Love, 2013; Purcell et al., 2010; Turner, Forstmann, et al., 2013), model-based neuroscientific analyses have been performed by way of a two-stage correlational procedure (Forstmann et al., 2008; Forstmann et al., 2010; Forstmann, Tittgemeyer, et al., 2011; Ho et al., 2012; Ho, Brown, & Serences, 2009; Liu & Pleskac, 2011; Philiastides, Ratcliff, & Sajda, 2006; Ratcliff, Philiastides, Sajda, 2009; Tosoni, Galati, Romani, & Corbetta, 2008; van Vugt, Simen, Nystrom, Holmes, & Cohen, 2012). In this procedure, the parameters of a cognitive model are first estimated by fitting the model to the behavioral data independently. Second, a neural signature of interest is extracted from the neural data alone, by way of either a statistical model or raw data-analytic techniques. Third, the behavioral model parameter estimates are correlated against the neural signature. Finally, significant correlations are used to substantiate claims of where the mechanisms assumed by the cognitive model are carried out in the brain.

The two-stage correlation procedure has greatly affected the emerging field of model-based cognitive neuroscience. For example, Forstmann and colleagues (Forstmann et al., 2008; Forstmann et al., 2010; Forstmann, Tittgemeyer, et al., 2011) have explored the contribution of the striatum and pre–supplementary motor areas (pre-SMA) to the response caution parameter in the linear ballistic accumulator (LBA; Brown & Heathcote, 2008) model. The response caution parameter represents the amount of remaining evidence an observer requires before eliciting a response. Forstmann and colleagues have studied how the response caution parameter relates to both neural activity (through fMRI; Forstmann et al., 2008) and neural structure (through diffusion-weighted imaging; Forstmann et al., 2010; Forstmann, Tittgemeyer, et al., 2011). Taken together, their efforts have brought forth a significant understanding of how the pre-SMA and striatum facilitate the flexible adjustment of response caution under a variety of

[1] Note that this does not characterize all cognitive neuroscientists—there are many researchers who rely heavily on cognitive models.

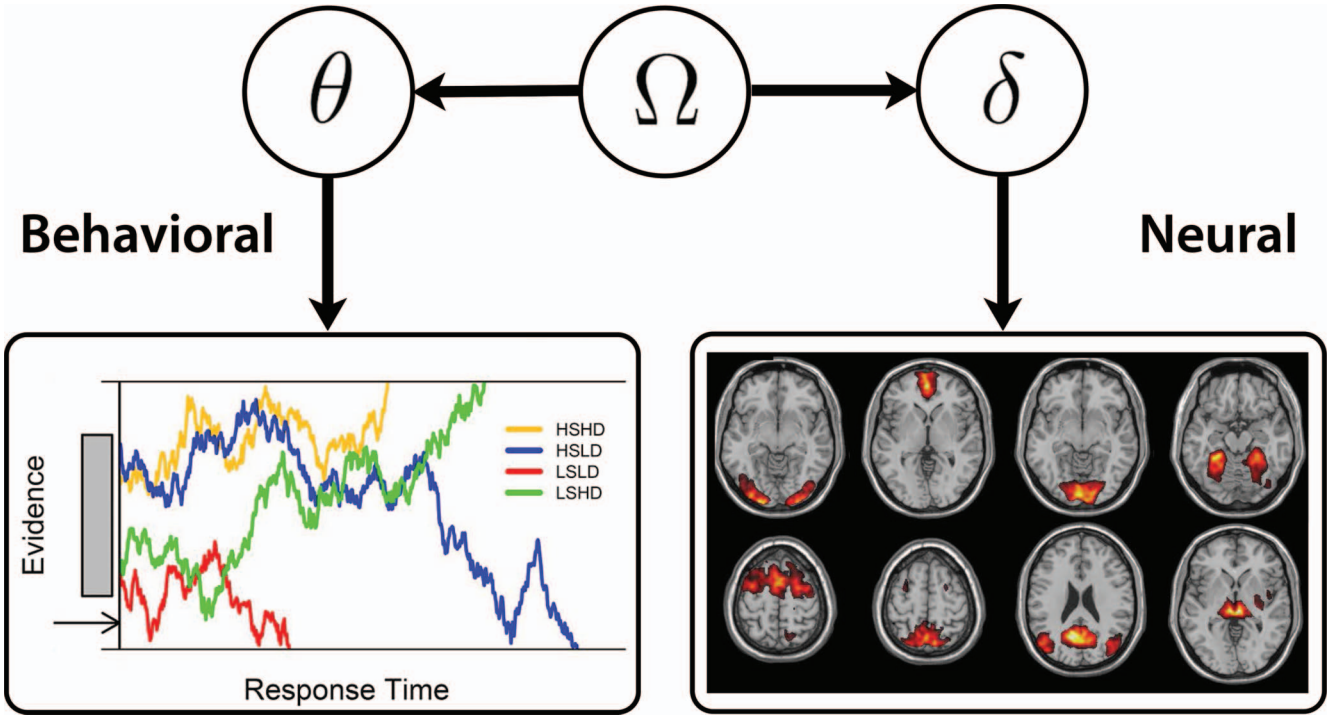

Figure 1. A graphical diagram of the neural drift diffusion model. The left side illustrates the standard DDM with parameters **θ** (i.e., the behavioral model) for four different types of responses within the model: high starting point and drift (HSHD; orange), low starting point and drift (LSLD; red), high starting point but low drift (HSLD; blue), and low starting point but high drift (LSHD; green). The right side illustrates activation patterns for eight regions of interest on a single trial (i.e., the neural model with parameters **δ**). The behavioral and neural models are conjoined by a hierarchical structure (middle) in which relationships between the mechanisms of cognitive models and statistical models of brain function are learned through the hyperparameters **Ω**.

speed emphasis instructions (e.g., emphasizing fast or accurate responses).

Most model-based cognitive neuroscience studies have focused on relating behavioral model parameters to neural activity aggregated across a set of trials. Conventionally basing an inference on aggregated data has produced many infamous misinterpretations and limitations (e.g., Heathcote, Brown, & Mewhort, 2000; Morey, Pratte, & Rouder, 2008). For example, Heathcote et al. (2000) showed how averaging across subjects produced a bias in favor of the power function in relating response times to number of practice trials. They found that when data were not aggregated across subjects, an exponential model produced a better fit to their data than the power function, which has important implications for psychological theory. To avoid the issues associated with aggregating data, van Maanen et al. (2011) examined the relationship between neural and behavioral measures on the single-trial level. To accomplish this, they employed a two-stage estimation procedure to obtain correlates of the single-trial parameters of the LBA model with the blood oxygen level-dependent (BOLD) signal. van Maanen et al. found that when subjects were told to respond quickly, the single-trial response caution parameter positively correlated with the BOLD signal in the pre-SMA and the dorsal anterior cingulate. However, when subjects were required to switch randomly between providing a fast or accurate response, the single-trial response caution parameter was positively correlated with the BOLD signal in the anterior cingulate proper. Although the approach by van Maanen et al. provided new insight into corticobasal ganglia functioning, their modeling efforts neglected an important source of parameter constraint. As we show later in this article, the two-stage correlation procedure they used is unable to exploit the constraint offered by the neural data. Therefore, the parameter estimates were noisy, possibly decreasing the strength of the reported correlational findings.

Although the two-stage correlation procedure has been useful in supporting various theories of cognitive processes, the procedure leaves much to be desired. First, the procedure generally aggregates across trial-to-trial information (but see van Maanen et al., 2011), which limits our understanding of how an observer's state of mind influences the behavioral data. Second, the procedure is not *statistically* reciprocal because the neural data cannot influence the parameter estimates of the behavioral model. Hence, the current state of model-based cognitive neuroscience neglects two important sources of significant constraint. In this article, our goal is to develop a framework that simultaneously obeys these constraints.

The Joint Modeling Framework

Turner, Forstmann, et al. (2013) proposed a new approach to model-based cognitive neuroscience that uses a hierarchical Bayesian framework to impose neurally based parameter constraints on behavioral models. Their framework, illustrated in

Figure 1, allows for a natural augmentation of cognitive models with neural data.[2] The left side of Figure 1 shows a behavioral model intended to capture the behavioral data (e.g., the drift diffusion model), and the right side displays a statistical model chosen to capture the neural data. To combine the two models, we assume a particular relationship between the behavioral model parameters and the statistical model parameters, regulated by the hyperparameters Ω (middle of Figure 1). The hierarchical structure allows for mutual constraint on all model parameters, captures multiple levels of effects (e.g., condition, subject, or trial effects), and provides a way of identifying more parameters than the number of data points on a given trial (see, e.g., Vandekerckhove et al., 2011). Finally, the modeling is carried out by examining the posterior distribution of the model parameters rather than conducting significance tests through correlation analyses, and so the framework inherits many advantageous features of Bayesian statistics. For example, the effects of interest can be examined without the need for Type I error corrections or the development of a priori hypotheses (O'Doherty, Hampton, & Kim, 2007).

Turner, Forstmann, et al. (2013) demonstrated the utility of their approach on simulated and experimental data by combining information across subjects. In their analysis, Turner et al. combined structural information measured by diffusion-weighted imaging and related this information to the parameters of the LBA (Brown & Heathcote, 2008) model. Because the neural measures were structural, they did not fluctuate from trial to trial, and as a consequence, the dynamics of the decision process could only be studied from one decision maker to another. Here, we advance this work by focusing the analysis on the single-trial level. In so doing, the new model allows us to better understand the dynamics of perceptual decision making by incorporating *functional* neurophysiological measures into our cognitive model. As we show below, our model enables us to identify brain regions associated with mechanisms assumed by our model, so that we can interpret neuroimaging data through the lens of a cognitive model.

The Neural Drift Diffusion Model

The neural drift diffusion model (NDDM) is a combination of a drift diffusion process for the behavioral data and a statistical model for the pattern of brain activity. Figure 1 shows a graphical diagram of the NDDM. On the left side, we use the drift diffusion model (DDM; Ratcliff, 1978) with parameters θ to model the joint distribution of choice and response time in a two-alternative forced-choice task. On the right side, we use a statistical model with parameters δ for the single-trial analysis of fMRI data (Eichele et al., 2008). The structural relationships between the pattern of brain activity and the cognitive model parameters are learned through the hyperparameters Ω, which define the hierarchical structure (Turner, Forstmann, et al., 2013).

Accounting for Behavioral Data

Sequential sampling models have provided a robust mathematical framework for understanding the cognitive processes underlying perceptual decision making (Ratcliff, 1978; Usher & McClelland, 2001). The fundamental structure of these models involves several parameters, which typically correspond to a response threshold, a nondecision component, a "drift" rate, and a bias parameter. Following the encoding of a stimulus, sequential sampling models assume that an observer accumulates evidence for each of a number of alternatives at a rate determined by the drift rate parameter. Once enough evidence has been accumulated for a particular alternative such that the evidence exceeds the response threshold, a response is triggered corresponding to the winning alternative.

Many neurophysiological findings have supported the notion of a sequential sampling process during perceptual decision making in nonhuman primates (e.g., Hanes & Schall, 1996; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Kiani, Hanks, & Shadlen, 2008). In humans, researchers have relied on the statistical relationship between measures obtained through noninvasive procedures (e.g., fMRI and EEG) and the latent mechanisms assumed by the cognitive models. One popular sequential sampling model is the DDM (Ratcliff, 1978). The DDM, and by extension the NDDM, identifies four different sources of variability in decision making: moment-to-moment variability within a single trial, trial-to-trial variability in the start point, nondecision time, and rate of evidence accumulation. The left bottom panel of Figure 1 shows an illustration of the DDM under four settings of the parameter values: a high starting point and high drift rate (HSHD; orange), a low starting point and low drift rate (LSLD; red), a high starting point but low drift rate (HSLD; blue), and a low starting point but high drift rate (LSHD; green). The left gray box represents the between-trial variability in the starting point, and the trajectories of the four model simulations show the within-trial variability in evidence accumulation. Each parameter regime corresponds to behavioral outcomes that are conceptually different. For example, the LSLD and HSLD examples both reach the bottom boundary (i.e., the incorrect decision) but do so at different times. In the LSLD regime, the model reaches the decision much faster than in the HSLD regime, because in the LSLD regime, the model does not need to overcome the initial bias toward the incorrect decision as in the HSLD regime. These two parameter settings produce the same response, albeit at different times, where one decision is primarily driven by an initial bias (i.e., the LSLD regime), and the other is more directly affected by the stimulus information (i.e., the HSLD regime). Figure 1 also shows how the within-trial variability in the model could potentially produce a response that is inconsistent with the stimulus information, which primarily influences the drift rate. For example, in the LSHD regime, the model is initially biased for the incorrect decision (i.e., the bottom threshold) and nearly reaches the bound early on. However, as time progresses, the model is able to recover from this initial bias and allows the stimulus information to drive the decision to the correct response.

Within the NDDM, we can separate the influence of (prestimulus) bias from the rate of stimulus information processing, allowing us to differentiate between competing explanations for how the data arise. For example, a fast correct decision might be the result of a particularly easy stimulus (e.g., a high drift rate), or it might just be a guess with a fortunate outcome (e.g., a high starting point). Distinguishing the various sources of variability in the

[2] Although the inclusion of neural data is particularly relevant here, the joint modeling framework is more general—any auxiliary data source can be inserted into the model.

decision process will be essential to understanding their relationship to neurophysiological measures. However, the inclusion of moment-to-moment variability within a single trial makes the calculation of the model's likelihood function difficult to evaluate. Here we use an efficient algorithm for calculating the likelihood of the DDM portion of our hierarchical model (Navarro & Fuss, 2009).

Using the DDM, Ratcliff and colleagues (2009) examined the correlations between single-trial EEG amplitude measures and condition-specific model parameters. Although the DDM does inherently have single-trial parameters, they were integrated out to facilitate parameter estimation. Here, we exploit the hierarchical structure of the NDDM to obtain estimates of the trial-to-trial parameters, an idea that is similar to other recent Bayesian treatments of the DDM (Vandekerckhove et al., 2011), but has the added benefit of combining neural measures to create a single, unified model of cognition (Turner, Forstmann, et al., 2013).

Accounting for Neural Data

Many fMRI findings have identified brain regions that seem to be related to sequential sampling processes (Mulder, van Maanen, & Forstmann, 2014). For example, Mulder, Wagenmakers, Ratcliff, Boekel, and Forstmann (2012) argue that the orbitofrontal cortex (OFC) is involved in generating decision bias, based on the correlation between OFC activation and DDM's bias parameter in various tasks. Analogous to the behavioral models, most of these studies have focused on grouped data, rather than single-trial fluctuations. However, single-trial parameter estimates of hemodynamic response functions can be obtained (Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; Eichele et al., 2008). The fluctuations in these parameters may be informative of predicting the fluctuations in behavior, and there are many ways in which single-trial parameters may be obtained. For example, Eichele et al. (2008) applied independent component analysis followed by deconvolution of the BOLD response to understand how brain activity preceding an error evolves from trial to trial. Using the same method, van Maanen et al. (2011) studied how the fluctuations in response caution (as measured with a single-trial sequential sampling model) are related to the trial-to-trial dynamics of the BOLD signal.

Technical Details

To explain the model more formally, we denote the threshold parameter as α, the single-trial starting point as $\boldsymbol{z} \in [0, \alpha]$, and the single-trial drift rate as $\boldsymbol{\xi}$.[3] For the present article, we ignore subject-specific differences and focus on trial-specific effects, although one could extend the model to capture these additional features of the data. Given this decision, we will drop the subject-specific subscript j from the exposition that follows. The single-trial parameter matrices (e.g., $\boldsymbol{\xi}$) are of length I, where I is the number of trials. We first reparameterize the starting point to be a proportion of the threshold, so that $\boldsymbol{w} = z/\alpha$, where $\boldsymbol{w} \in [0, 1]$. Using this notation, the probability density function for the first passage time distribution for the lower boundary is given by

$$f(t|\alpha, w_i, \xi_i) = \frac{\pi}{\alpha^2}\exp\left(-\xi_i \alpha w_i - \frac{\xi_i^2 t}{2}\right) \times \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2\pi^2 t}{2\alpha^2}\right)\sin(k\pi w_i) \quad (1)$$

(Feller, 1968; Navarro & Fuss, 2009). To obtain the probability density at the upper boundary, we replace ξ_i with $-\xi_i$ and w_i with $1 - w_i$ in Equation 1. We incorporate a nondecision time parameter τ by replacing t in Equation 1 with $(t - \tau)$. If we arbitrarily define $c = 1$ for the Wiener diffusion process that absorbs at the lower bound and $c = 2$ otherwise, the joint probability density function for a given response c at time t is given by

$$\text{Diffusion}(c, t|\alpha, \xi_i, w_i, \tau) = f\left(t - \tau|\alpha, c - 1 + (-1)^{c-1}w_i, (-1)^{c-1}\xi_i\right). \quad (2)$$

To obtain parameters with infinite support, we transform the single-trial start points w by a logistic transformation so that

$$\boldsymbol{\omega} = \text{logit}(\boldsymbol{w}) = \log\left(\frac{\boldsymbol{w}}{1 - \boldsymbol{w}}\right).$$

One could also assume that the nondecision time parameter τ fluctuated from trial to trial. However, in the current implementation of the NDDM, we chose against adding this additional layer of complexity for reasons of parsimony, a decision that has been supported by other researchers (e.g., Brown & Heathcote, 2008; Usher & McClelland, 2001). Furthermore, in the NDDM, the nondecision time parameters capture effects that are not cognitively interesting, and so we decided to save investigation of the neural basis of this mechanism for future research.

Multiple methods are available for estimating trial-to-trial variability in the neural data (e.g., Eichele et al., 2008; Mumford, Turner, Ashby, & Poldrack, 2012; Waldorp, Christoffels, & van de Ven, 2011). For the data reported here, we used independent component analysis (ICA; Calhoun, Adali, Pearlson, & Pekar, 2001; see Methods section for details). Given a set of M neural sources, we denote the single-trial activation of the mth source (i.e., the estimate of the beta weight) on the ith trial as $\beta_{i,m}$.

To create a unified model of behavioral and neural data, we assume that the single-trial drift rates ξ_i, single-trial starting points ω_i, and the single-trial activation parameters for the M sources $\beta_{i,1:M}$ come from a common distribution, so that

$$(\xi_i, \omega_i, \beta_{i,1:M}) \sim \mathcal{MVN}(\boldsymbol{\phi}, \boldsymbol{\Sigma}),$$

where $\mathcal{MVN}(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean vector $\boldsymbol{\phi}$ and variance covariance matrix $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Omega} = \{\boldsymbol{\phi}, \boldsymbol{\Sigma}\}$ from Figure 1). The hypermean vector $\boldsymbol{\phi}$ is of length $(2 + M)$, whereas the variance-covariance matrix $\boldsymbol{\Sigma}$ is of dimension $(2 + M) \times (2 + M)$. Here the "2" corresponds to the two single-trial parameters ω and ξ.

[3] A word on notation is in order here. We use boldface fonts to represent entities containing more than one element and represent scalars in regular font. Sometimes it will be convenient to refer only to a portion of a matrix, and we do this by indicating the range of elements within a subscript. For example, the matrix ξ might have dimensions J by I, but we refer to the jth row as $\xi_{j,1:I}$.

The hypermean vector $\boldsymbol{\phi}$ represents the mean of the joint vector $[\boldsymbol{\theta}, \boldsymbol{\delta}]$ whereas $\boldsymbol{\Sigma}$ represents the variance covariance matrix. For ease of interpretation, we can partition the matrix $\boldsymbol{\Sigma}$ by combining the behavioral model variance-covariance matrix η^2 (of dimension $[2 \times 2]$) with the neural model activation variance-covariance matrix σ^2 (of dimension $[M \times M]$), so that

$$\Sigma = \begin{bmatrix} \boldsymbol{\eta^2} & \boldsymbol{\eta\rho\sigma} \\ (\boldsymbol{\eta\rho\sigma})^{\mathrm{T}} & \boldsymbol{\sigma}^2 \end{bmatrix}, \quad (3)$$

where

$$\boldsymbol{\eta\rho\sigma} = \begin{bmatrix} \eta_{1,1}\rho_{1,1}\sigma_{1,1} & \eta_{1,1}\rho_{1,2}\sigma_{2,2} & \cdots & \eta_{1,1}\rho_{1,M}\sigma_{M,M} \\ \eta_{2,2}\rho_{2,1}\sigma_{1,1} & \eta_{2,2}\rho_{2,2}\sigma_{2,2} & \cdots & \eta_{2,2}\rho_{2,M}\sigma_{M,M} \end{bmatrix},$$

a matrix of dimension $(2 \times M)$. The matrix ρ (of dimension $[2 \times M]$) is a correlation matrix that assesses the relationship between each region of interest (ROI) and each cognitive model parameter. Hence, if for example $\rho_{1,1}$ is estimated to be near zero, the interpretation is that the first behavioral model parameter (i.e., the drift rate) is unrelated to the first neural source. This feature of our model allows us to enforce a priori considerations, such that if a particular neural source is known to be uncorrelated with a particular behavioral model parameter, one can simply constrain $\rho_{1,1} = 0$. Because we have no a priori beliefs, we will freely estimate all parameters in $\boldsymbol{\rho}$.

Importantly, the parameters of the NDDM are conjoined with the neural measures in our data on every trial, and so the variability in the BOLD activation is linked to the trial-to-trial variability within the model. Our modeling approach allows us to examine brain-behavior patterns in several novel ways. First, it allows us to explore the data on the basis of both choice and response time measures. Second, the NDDM possesses parameters that capture both single-trial and subject-specific effects. Third, the modeling framework enforces neural constraints on the behavioral model parameters and vice versa, which may have important consequences for model development (cf. Jones & Dzhafarov, 2014; Turner, 2013). Finally, by using our model-based cognitive neuroscience approach, we make an explicit link to the computational theory of sequential sampling. In so doing, we can provide a mechanistic view of the BOLD activity by way of cognitively meaningful constructs such as the rate of information processing or a prestimulus bias.

In the next section, we present the results of a simulation study designed to examine the contributions of linking the single-trial parameters with auxiliary single-trial data (i.e., neural data). In particular, we show how a model that integrates neural and behavioral data on a single-trial basis can outperform a model that only takes advantage of the behavioral data.

Simulation

The main objective of our simulation is to investigate the predictions from two models: the first model—the DDM—only takes into account the behavioral data, whereas the second model—the NDDM—integrates both behavioral and neural data in the way described earlier. Because the DDM only accounts for behavioral data, it can only generate predictions for future behavioral data based on the information it learns from past behavioral patterns. On the other hand, the NDDM takes advantage of past behavioral performance as well as current neurophysiological information on every trial. Therefore, the advantages in predictive performance provided by the NDDM will be directly proportional to the magnitude of the statistical association between the neural signature and the behavioral model parameters (Turner, 2013).

Perhaps the best way to evaluate the relative contributions of the two models is to examine the accuracy of their predictions for future data (i.e., data not used in the fitting process). To do so, we first generated data from a joint model for a single subject (and so we will drop the subject index in all parameters below) that linked a single-trial neural signature to the single-trial parameters in the behavioral model—specifically the drift rate $\boldsymbol{\xi}$ and starting point $\boldsymbol{\omega}$ matrices. For the neural signature, we assumed the presence of 34 brain regions or "sources" whose single-trial activation parameter $\boldsymbol{\beta}$ fluctuated from trial to trial.[4] Specifically, we assumed

$$(\xi_i, \omega_i, \beta_{i,1:M}) \sim \mathcal{MVN}(\boldsymbol{\phi}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ were chosen to be similar to the data in our experiment reported below. Once $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ had been established, they were used to randomly generate the parameter vectors $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, and $\boldsymbol{\beta}$. We set the threshold parameter equal to $\alpha = 2$ and the nondecision time parameter equal to $\tau = 0.1$, which are settings of the parameters that produce data reasonably close to the experimental data reported below.

Particularly relevant within Σ is the correlation matrix $\boldsymbol{\rho}$. Figure 2 shows a histogram of the elements within $\boldsymbol{\rho}_{1,1:M}$ (left panel) and $\boldsymbol{\rho}_{2,1:M}$ (right panel). The figure shows that many of the elements are near zero and, as a consequence, enforce minimal constraint on the behavioral model parameters (Turner, 2013). Of the 68 elements of $\boldsymbol{\rho}$, only 14 elements have absolute values greater than 0.2. Although this may seem like an inconsequential number, we see later how only a few nonzero elements of $\boldsymbol{\rho}$ can be used to heavily constrain the parameters of the behavioral model, ultimately leading to better predictions by the NDDM relative to the DDM.

Using the parameter settings described above, we generated 400 choice response times and source activations from the NDDM (i.e., the observations were all independent and identically distributed). Next, both the NDDM and the DDM were trained on the first 200 data points (i.e., the "training data"). Finally, each model generated predictions about the remaining 200 data points (i.e., the "test data"). Given the nature of the models, a few model details differ between the NDDM and DDM, which we now discuss.

Model Details

The NDDM. The details of the NDDM are outlined above, with the exception of the particular choices we made about prior distributions. Here, we specified noninformative priors for α and τ such that

$$\tau \sim U[0, 2]$$

$$\alpha \sim U[0, 10],$$

where $U[a, b]$ denotes the uniform distribution on the interval $[a, b]$. We specified a joint prior on the hypermean parameter vector $\boldsymbol{\phi}$ and hyper-variance-covariance matrix $\boldsymbol{\Sigma}$ so that

[4] We chose 34 neural sources to match the number of sources used in our experimental data below.

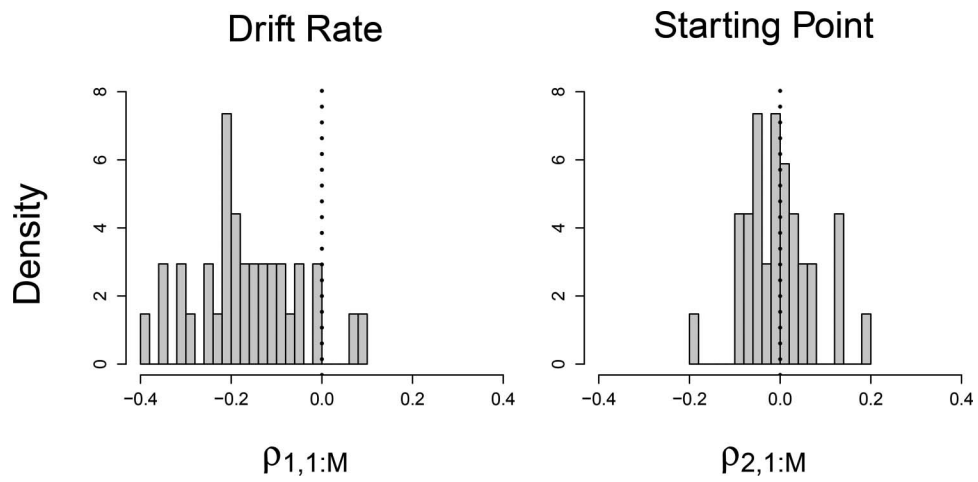

Figure 2. Histogram of the correlation matrix $\boldsymbol{\rho}$ used in the simulation and estimated from the experimental data. The left panel shows the elements of the matrix $\boldsymbol{\rho}_{1,1:M}$ corresponding to the associations between the 34 neural sources and the drift rate parameter, whereas the right panel shows the elements of $\boldsymbol{\rho}_{2,1:M}$ corresponding to the starting point parameter.

$$p(\boldsymbol{\phi}, \boldsymbol{\Sigma}) = p(\boldsymbol{\phi}|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\phi}|\Sigma \sim \mathcal{MVN}(\boldsymbol{\mu}_0, \boldsymbol{s}_0^{-1}\boldsymbol{\Sigma}),$$

and

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Phi}, d_0),$$

where $\mathcal{W}^{-1}$ ($\boldsymbol{a}$, b) denotes the inverse Wishart distribution with dispersion matrix $\boldsymbol{a}$ and degrees of freedom b. We set d_0 equal to the number of sources plus the number of single-trial model parameters plus two (i.e., $d_0 = 34 + 2 + 2 = 38$), $s_0 = 1/10$, and $\boldsymbol{\mu}_0$ is a vector containing 36 zeros. These choices were made to establish a conjugate relationship between the prior and posterior, so that analytic expressions could be derived for the conditional distributions of $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$, while still specifying uninformative priors.[5]

The DDM. We chose to maintain as much consistency across the joint and behavioral models as possible, which resulted only in changes in the prior specifications. The only difference between the models was that the behavioral model ignored the neural data. Because the number of neural sources is zero, we set $d_0 = 4$. Furthermore, the neural model parameters $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\sigma}$ were removed from the model. Given these specifications, the behavioral model we employed is slightly more flexible than the classic DDM (Ratcliff, 1978), because it contains a parameter that models the correlation between the single-trial starting point and drift rate. As a consequence of this additional flexibility, the DDM we use here performed slightly better than the classic DDM in some preliminary simulation results.

Results

Because we are first testing the model in a simulation study, we have the advantage of knowing the true values of the model parameters. For now, we will compare the models on their ability to predict the single-trial model parameters rather than actual behavioral data. The reason for this decision is purely for illustrative purposes: it is difficult to visually compare the choice response time predictions of the two models for all of the test data in a Bayesian setting. However, in the experimental data reported below, because we do not know what the true model parameters are, we will be unable to compare the models on that basis and will resort to statistical summaries of model predictions for behavioral data instead.

Figure 3 shows the predictions for the single-trial parameters of the NDDM (left column; green) and the DDM (right column; red) against the test data (black lines). The top row of Figure 3 corresponds to the drift rates, whereas the bottom row corresponds to the starting points. In both rows, the true model parameters for the test data have been sorted in increasing order to better facilitate a comparison of the models' predictions. In each panel, the maximum a posteriori (MAP) prediction is shown as the dark correspondingly colored line, and the 95% credible sets are shown as the light-shaded region.

Because the DDM neglects the neural signature, the model can only generate predictions for the single-trial parameters based on previous experience with the task (i.e., the training data). Although these are optimal predictions in a statistical sense, they are—as the right panel of Figure 3 demonstrates—completely insensitive to trial-to-trial fluctuations in the model parameters. On the other hand, the joint model takes advantage of (a) the trial-to-trial fluctuations in the neural signature and (b) the association between source activation and drift rate. From these two features of the data, the joint model makes appreciably better predictions for the single-trial drift rates relative to the behavioral model. The correlation between the true and predicted drift rates for the NDDM was 0.273, $t(198) = 3.99$, $p < .01$, whereas the correlation for the

[5] We remained agnostic about the specification of the priors because this is the first time our model has been fit to data.

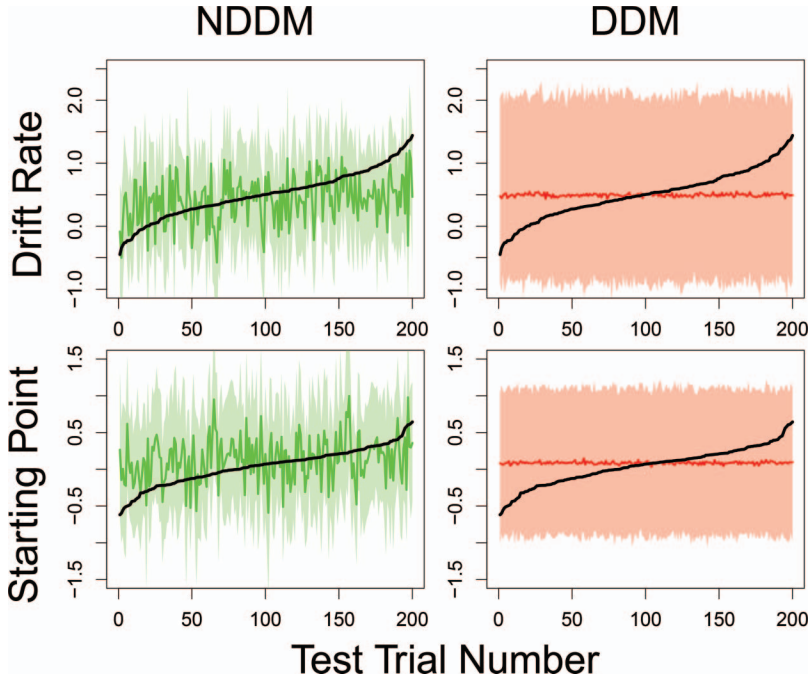

Figure 3. Model predictions for single-trial parameters. The left panels show the predictions for the neural drift diffusion model (NDDM), whereas the right panels show the predictions for the drift diffusion model (DDM). The top row corresponds to the single-trial drift rates, whereas the bottom row corresponds to the single-trial starting points. The 95% credible set is shown as the shaded region (green for the NDDM and red for the DDM), the maximum a posteriori prediction is shown in the corresponding color, and the true single-trial parameters are shown in black. The single-trial parameters have been sorted in increasing order to better illustrate the differences in model predictions.

DDM was 0.072, $t(198) = 1.01$, $p = .31$. The correlation between the true and predicted starting points for the NDDM was 0.205, $t(198) = 2.95$, $p < .01$, whereas the correlation for the DDM was 0.057, $t(198) = 0.80$, $p = .43$.

It is important to note that the correlation between the model parameters and the neural data is in itself not sufficient for allowing the NDDM to outperform the DDM on the prediction task. Because the NDDM has several more parameters relative to the DDM, the strength of the association between the model parameters and the neural data must be strong enough to warrant the additional flexibility of the NDDM. For example, if there was no correlation between the model parameters and the neural data (i.e., all off-diagonal elements of $\mathbf{\Sigma} = 0$), the predictions of the NDDM would on average match the predictions of the DDM, but the predictions of the NDDM would be more variable (the shaded green area corresponding to the NDDM in Figure 3 would be larger than the shaded red area corresponding to the DDM) relative to the DDM as a result of the additional parameter uncertainty in the model. This is important because the information in the right panel of Figure 3 is what would be used to generate predictions in the two-stage correlation procedure.

The primary question of interest in this article is whether neural data can be used in a generative manner to enhance our understanding of the mind from a mechanistic point of view. Up to this point, we have developed a model of behavioral and neural data and shown that in at least one situation, the model can outperform the core behavioral model (also see Turner, 2013; Turner, Forstmann, et al., 2013). Although we chose values of the parameters to produce data that were reasonably close to experimental data, it remains unclear whether the NDDM is an enhancement of the DDM for truly experimental data. In the next section, we fit the models to experimental data to further test the merits of the NDDM relative to the DDM.

Experiment

To test our hypothesis that there exists an association between single-trial model parameters and trial-to-trial fluctuations in the neural signal, we rely on data reported in (van Maanen et al., 2011), which collected response choice, response times, and the prestimulus BOLD signal for every trial under two speed emphasis conditions. In the accuracy condition, subjects were instructed to respond as accurately as possible, whereas in the speed condition, subjects were instructed to respond as quickly as possible. The experiment used a moving dots task where subjects were asked to decide whether a cloud of semirandomly moving dots appeared to move to the left or to the right. Subjects indicated their response by pressing one of two spatially compatible buttons with either their left or right index finger. Before each decision trial, subjects were instructed whether to respond quickly (the speed condition) or accurately (the accuracy condition). Following the trial, subjects were provided feedback about their performance. We implemented the speed-accuracy manipulation to verify that our modeling results were consistent with prior work on response caution and to facilitate estimation of the model parameters. However, a block-

type difficulty manipulation was not necessary because, as we show, the variability inherent in the experimental stimulus generation procedure alone allowed for a more informative (i.e., continuous) difficulty manipulation.

Participants

Seventeen participants (seven female; mean age = 23.1 years, *SD* = 3.1 years) gave informed consent and participated in this experiment. All participants had normal or corrected-to-normal vision, and none had a history of neurological, major medical, or psychiatric disorders. All participants were right-handed, as confirmed by the Edinburgh Inventory (Oldfield, 1971). The experimental standards were approved by the local ethics committee of the University of Leipzig. Data were handled anonymously.

Stimuli

Participants performed a two-alternative forced-choice random dot motion task. On each trial, participants were asked to decide whether a cloud of three-pixel-wide dots appeared to move to the left or the right. The cloud consisted of 120 dots, of which 60 moved coherently and 60 moved randomly. "Coherence" was achieved by moving the coherent dots one pixel in the target direction from frame to frame. The remaining dots were relocated randomly. On the subsequent frame, the coherent dots were moved randomly, and the random dots were now treated as coherent, such that the appearance of motion was determined by all dots, and participants could not focus on one dot to make a correct inference. The diameter of the entire cloud circle was 250 pixels. In this circle, pixels were uniformly distributed. Importantly, the variability in the stimulus-generation process combined with our single-trial analysis allows us to better investigate how stimulus difficulty interacts with the decision process relative to other, block-type manipulations (Liu & Pleskac, 2011; Ratcliff et al., 2009).

fMRI Data Acquisition and Analysis

Each trial lasted 10 s and started with a jittered temporal interval between 0 and 1,500 ms. Then a cue was presented in the middle of the screen for 2,000 ms, indicating whether the trial was speed stressed or accuracy stressed. Cue presentation was followed by a jittered interval between 0 and 3,000 ms in steps of 1,000 ms. The dot cloud was presented for 1,500 ms and followed by feedback. On the speed-stressed trials, participants were required to respond within 400 ms after stimulus onset. On the accuracy-stressed trials, participants were required to respond within 1,000 ms.

The experiment was conducted in a 3T scanner (Philips) while whole-brain fMRI was obtained. The fMRI data were acquired prior to the onset of the condition cue. Thirty axial slices were acquired (222 × 222 mm field of view, 96 × 96 in-plane resolution, 3-mm slice thickness, 0.3-mm slice spacing) parallel to the anterior commissure–posterior commissure plane. We used a single-shot, gradient-recalled echo planar imaging sequence (repetition time 2,000 ms, echo time 28 ms, 90-degree flip angle, transversal orientation). Prior to the functional runs, a 3D T1 scan was acquired (T1, 25 × 25 cm field of view, 256 × 256 in-plane resolution, 182 slices, slice thickness 1.2, repetition time 9.69, echo time 4.6, sagittal orientation).

As previously reported (van Maanen et al., 2011), the independent components were obtained through group spatial ICA (Calhoun et al., 2001). For each individual separately, the preprocessed fMRI data were prewhitened and reduced via temporal principal component analysis (PCA) to 60 components. Then, group-level aggregate data were generated by concatenating and reducing individual principal components in a second PCA step. Infomax ICA (Bell & Sejnowski, 1995) was performed in this set with a model order of 60 components (Kiviniemi et al., 2009). To estimate robust components, we used independent component analysis software package (Himberg, Hyvarinen, & Esposito, 2004)—that is, the decomposition was performed 100 times with random initial conditions—and identified centroids with a canonical correlation-based clustering. Individual independent component maps and time courses were reconstructed by multiplying the corresponding data with the respective portions of the estimated demixing matrix. The set of 60 components was further reduced by excluding components that were associated with artifacts, had a cluster extent of fewer than 26 contiguous voxels, and had uncorrected *t* statistics of $t_{19} < 5$. Consequently, 34 ICs were considered in the main analyses (we maintained the original numbering of 1–60). Although doing the preprocessing (i.e., the ICA) was not actually necessary, it helped us to reduce the dimensionality of the problem because instead of modeling activity in a (large) number of voxels, we could instead focus our efforts on modeling activity in a small set of important brain regions.

To obtain single-trial estimates of the hemodynamic response (HR) amplitudes, we used the method reported in Eichele et al. (2008) and Danielmeier et al. (2011). For each participant and component separately, the empirical event-related HR was deconvolved by forming the convolution matrix of all trial onsets with a length of 20 s and multiplying the Moore-Penrose pseudoinverse of this matrix with the filtered and normalized component time course. Single-trial amplitudes were recovered by fitting a design matrix containing separate predictors for each trial onset convolved with the estimated HR onto the IC time course, estimating the scaling coefficients ($\boldsymbol{\beta}$) by using multiple linear regression. Further details on the experimental design and the neural analysis can be obtained from van Maanen et al. (2011).

Model Details

All choices for the models were equivalent to those made in the simulation study. For these data, we allow the threshold parameter to vary across speed emphasis conditions, which is consistent with numerous sequential sampling model accounts of the speed-accuracy trade-off (e.g., Boehm, Van Maanen, Forstmann, & Van Rijn, 2014; Mulder et al., 2013; Turner, Sederberg, Brown, & Steyvers, 2013; Winkel et al., 2012). We use a vector of response threshold parameters $\alpha = \{\alpha^{(1)},\alpha^{(2)}\}$ so that $\alpha^{(1)}$ and $\alpha^{(2)}$ are used for the accuracy and speed conditions of the experiment, respectively. Similarly, we use a vector of nondecision time parameters $\tau = \{\tau^{(1)},\tau^{(2)}\}$ for the same respective conditions.

We denote the reaction time (RT) for the *j*th subject on the *i*th trial as $RT_{j,i}$, the response choice as $RE_{j,i}$, and the speed emphasis condition as $C_{j,i}$. Thus, the behavioral data for Subject *j* are $\mathbf{B}_{j,1:I} = \{\mathbf{RE}_{j,1:I}, \mathbf{RT}_{j,1:I}, \mathbf{C}_{j,1:I}\}$. Given the priors listed in the simulation section, and denoting the neural data for the *j*th subject as $N_{j,1:I,1:M}$, the full joint posterior distribution for the model parameters is given by

$$p(\Xi|\mathbf{B},\mathbf{N}) \propto \prod_{j=1}^{J}\prod_{i=1}^{I} \text{Diffusion}\left(RE_{j,i}, RT_{j,i}|\alpha^{(C_{j,i})}, \omega_i, \tau, \xi_i\right) \times \prod_{i=1}^{I}\left[\prod_{j=1}^{J}\prod_{m=1}^{M} p(N_{j,i,m}|\beta_{i,m})p\left(\xi_i, \omega_i, \beta_{i,m}|\boldsymbol{\phi}, \boldsymbol{\Sigma}\right)\right] \times \prod_{k=1}^{2} p\left(\alpha^{(k)}\right)p\left(\tau^{(k)}\right)p(\boldsymbol{\phi}\,|\,\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}),$$

where $p(N_{j,i,m}|\beta_{i,m})$ was defined in Eichele et al. (2008) and

$$\Xi = \{\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\xi}_{1:I}, \boldsymbol{\omega}_{1:I}, \boldsymbol{\beta}_{1:I,1:M}, \boldsymbol{\phi}, \boldsymbol{\Sigma}\}.$$

Results

We used a combination of Gibbs sampling (Gelman, Carlin, Stern, & Rubin, 2004) and differential evolution with Markov chain Monte Carlo (ter Braak, 2006; Turner, Sederberg, et al., 2013) to fit the model to the data. We ran the algorithm for 15,000 iterations with 24 chains following a burn-in period of 5,000 samples. We thinned the chains by retaining every third iteration, resulting in 120,000 samples of the joint posterior distribution. Standard techniques were used to assess convergence (Plummer, Best, Cowles, & Vines, 2006).

Evaluating Model Fit

Before we begin interpreting posterior distributions, it is important to verify that the model fits adequately to the data. For the NDDM, this assessment involves verifying that the model can (a) recover the pattern of source activations present in the neural data and (b) produce behavioral data that match the observed behavioral data reasonably well. To verify objective (a), we compared the statistics of the raw neural data (e.g., the source means and correlation matrix) to the estimated posterior distributions of model parameters (e.g., $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$). Between the two comparisons, the more difficult task is in the comparison of Σ, so we will only present those results here.[6] Figure 4a and 4b shows the assessment of model fit to the neural data. Figure 4a plots the correlation matrix of the 34 sources estimated from the raw ICA data against the predicted relationships generated from the model. The left panel of Figure 4b shows the correlation matrices for the raw ICA data, whereas the right panel shows the MAP estimate for each element of the corresponding correlation matrix to the variance-covariance matrix $\boldsymbol{\sigma}^2$ predicted by the NDDM (see Equation 3). Recall that the variance-covariance matrix $\boldsymbol{\sigma}^2$ is the partition of $\boldsymbol{\Sigma}$ that exclusively handles the interrelationships between the neural sources. The correlation matrices are symmetric, so, for example, the element in the 30th row, second column is equal to the element in the 33rd row, fifth column. Each element of the matrix is color coded according to the legend on the far-right panel, where high correlations are shown in red (i.e., correlations of 0.5) and low correlations are shown in blue (i.e., correlations of −0.2). The diagonal elements of this matrix are all equal to 1.0 (shown in black), but we have removed them for illustrative purposes. Adequate recovery of $\boldsymbol{\sigma}^2$ implies that the NDDM is capturing the patterns of source activity observed in the data, for reasons that are both functional (e.g., connectivity) and spatial (e.g., proximity). Figure 4a and 4b shows that we have accurately recovered the pattern of source activations in our data.

To evaluate the model's fit to the behavioral data, we examined the posterior predictive distribution (PPD) of the choice response time distributions. The PPD serves as a generalization of the information obtained in the empirical data to new, hypothetical data that might have been observed had more trials been obtained in the experiment. The PPD provides a statistically coherent way to simultaneously form a quantification of uncertainty and establish a "best" estimate for the predicted model parameters, based on the data that were observed. To generate the PPD, we focused on the behavioral model parameters $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$, rather than the single-trial parameters ξ and ω, so that the model predictions were generalized to hypothetical data that could have been observed had more data been collected. Aggregating across the parameter space in this way will produce model predictions that are inherently more variable and, generally speaking, less accurate. Recall that the parameters $\boldsymbol{\theta}$ are the subset of the hyperparameter ϕ that are exclusive to the behavioral model (i.e., $\boldsymbol{\theta}$ is the hypermean of ξ and ω). To generate the most likely PPD, we obtained the MAP estimate for each model parameter and plotted the corresponding defective distributions for each stimulus by speed emphasis condition. Figure 4c to 4f shows the choice response time distributions from the empirical data (histograms) along with the MAP prediction from the model (black lines) for each combination of stimulus and speed emphasis condition: (c, d) rightward stimulus presentations, (e, f) leftward stimulus presentations, (c, e) accuracy emphasis condition, and (d, f) speed emphasis condition. In each panel, the defective response time distributions are shown for both the "leftward" (left) and "rightward" (right) responses, meaning that the probability of each response alternative from the model can be evaluated by comparing the heights of the two distributions. Figure 4c to 4f shows that the PPD closely resembles the basic form of the empirical data, demonstrating that the model is accurately fitting the data. However, a skeptical reader may wonder why the model slightly misses some aspects of the data. Part of this effect is due to a phenomenon known as *shrinkage*. This means that the model is forced to capture both the neural and behavioral data, where the neural data contain information about the activations in 34 ROIs and the behavioral data contain only two observations in the form of a choice response time pair. Because the neural data are more numerous, the model places more emphasis on accurately fitting the neural data relative to the behavioral data, and so if any misfit should occur, it will be more likely to occur on the least informative data measures (i.e., the behavioral data in this case).

Regarding the additional parameters in the model, the posterior distribution of the threshold parameter for the accuracy condition $\alpha^{(1)}$ had a median of 1.585 with a 95% credible set of (1.521, 1.658), whereas the threshold parameter for the speed condition $\alpha^{(2)}$ had a median of 1.372 with a 95% credible set of (1.323, 1.426). These parameters vary in a way that is predicted from the experimental manipulation—namely, that the response threshold should decrease as the speed of a response is emphasized over the accuracy of that response. The posterior distribution of the nondecision time parameter in the accuracy condition $\tau^{(1)}$ had a median of 0.299 with a 95% credible set of (0.291, 0.305), whereas the nondecision time parameter in the speed condition $\tau^{(2)}$ had a median of 0.057 with a 95% credible set of (0.053, 0.607). These parameters also vary in a way that is predicted from the experimental manipulation.

[6] Our results showed that $\boldsymbol{\phi}$ was recovered accurately compared to the empirical means of the raw neural data.

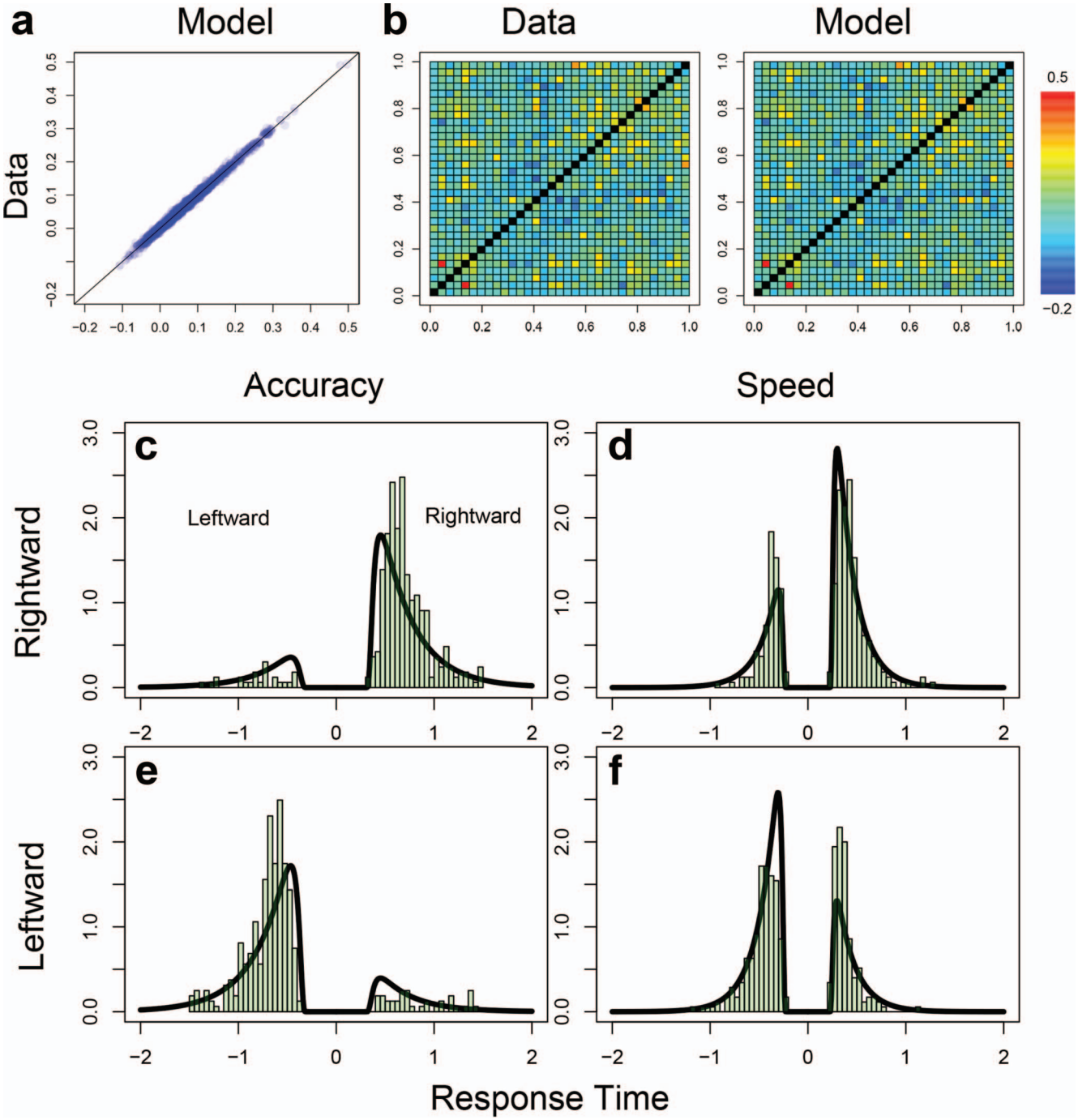

Figure 4. Evaluation of model fit. The top panels (a, b) show the model fit to the neural data. Specifically, Panel a plots the estimated region-to-region correlation matrix obtained from the raw data against the predicted relationship from the model. Panel b shows the correlation matrices estimated from the data (left) and the predicted relationships generated from the model (right). The bottom panels (c–f) show the model predictions (black lines) against the raw data (histograms) for the choice response time distributions in each of the various conditions in the experiment: (c, d) rightward stimulus presentations, (e, f) leftward stimulus presentations, (c, e) accuracy emphasis condition, and (d, f) speed emphasis condition. In each panel (c–f), defective response time distributions are shown for both the "leftward" (left) and "rightward" (right) responses.

Once assured that a reasonably accurate model fit had been obtained, we examined the relationship between patterns of BOLD activity and the parameters of the cognitive model. To do this, we first generated 10,000 samples from the estimated $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ parameters to form a more stable representation of the association between BOLD activity and parameters of the cognitive model.[7] We examined the resulting PPD of ($\boldsymbol{\theta}$, $\boldsymbol{\delta}$) to determine the extent to which the prestimulus BOLD activity in each ROI was predictive of the single-trial mechanisms used by the NDDM. ROIs were defined from the independent components extracted from our data and were not defined a priori (for more details, see Eichele et al., 2008).

Once the PPD was generated, we defined regions within the PPD that corresponded to psychologically interpretable constructs—namely, the rate of stimulus information processing, bias, and response efficiency. We defined efficiency as the ability to make fast, yet unbiased, decisions. The regions allowed us to generate psychologically meaningful constructs on the basis of the parameters in the model. The regions were chosen based on how the constructs mapped onto the mechanisms assumed by the NDDM and specifically the range of parameter values corresponding to these constructs. Within the NDDM, the rate of stimulus information processing corresponds to the drift rate, and the degree of bias corresponds to the starting point. Figure 5a, 5d, and 5g shows the joint PPD of drift rate and starting point. The rows of Figure 5 correspond to the three construct analyses we performed and will discuss in the sections that follow: drift rate (top), starting point (middle), and efficiency (bottom). Each colored region in Figure 5a, 5d, and 5g corresponds to a separate type of behavioral response pattern. Although the absolute locations of each region were defined after the model was fit to the data, the relative locations were defined a priori on the basis of cognitively meaningful constructs. For example, a "high" drift rate was defined to contain only relatively large values of drift, and "high" bias was defined to contain only starting points that were much closer to one response threshold than the other. Figure 5b, 5e, and 5h shows the predicted choice response time distributions for each correspondingly colored region in the joint posterior distribution under accuracy emphasis instruction, whereas Figure 5c, 5f, and 5i shows the choice response time distributions under speed emphasis instruction. The figure shows only the model's predictions for stimuli with rightward directional motion because the predictions for leftward motion stimuli were mirror images of the rightward motion stimuli.[8] In each of the response time distribution plots, distributions corresponding to the correct decision are shown on the left (i.e., negative values), whereas distributions corresponding to the incorrect decision are shown on the right. In each panel, the model's prediction for the probability of a correct response is represented as the density of the correct response time distribution relative to the incorrect response time distribution. In the sections that follow, we further investigate how activity in the brain relates to these three cognitive constructs in turn. Specifically, in the section entitled "Drift Rate Region Analysis," we examine the pattern of brain activations predicted by the model under high and low drift rate trials. Next, in the section entitled "Starting Point Region Analysis," we investigate the pattern of brain activations present during high- and low-bias trials, which is determined by the degree to which the starting point deviates from the point of equal preference for each alternative. Finally, in the section entitled "Efficiency Region Analysis," we investigate a new construct defined as the degree to which observers are able to integrate stimulus information in a way that is unbiased. In addition, in the drift rate region analysis, we explain our results in terms of the default mode network, which has been shown to be active when a subject is engaged in off-task behavior.

Drift Rate Region Analysis

The mechanism that most directly corresponds to the rate of stimulus information accumulation in the NDDM is the drift rate. Because the BOLD activity is obtained prior to the stimulus presentation, any significant relationship between the BOLD response and the drift rate is indicative that the prestimulus state of the mind is predictive of task behavior. Hence, our first region analysis focused on the drift rate and fits relation to the prestimulus BOLD signal. For this analysis, we picked two regions in the parameter space: one region corresponding to high drift rates (i.e., ξ, with an interval of $\pm$ 0.2) and one region corresponding to low drift rates (i.e., ξ, with an interval of $\pm$ 0.2). Figure 5a shows the two drift rate regions we selected, where the green region corresponds to the high drift rate region and the red region corresponds to the low drift rate region. As for the starting point parameter, we defined our regions to include a large range of values for the start point so that the effects of bias on the decision could be fully integrated out, isolating the contribution of drift rate in the decision process. To generate predictions for the behavioral data from the model, we generated a choice response time distribution by randomly selecting values for the drift rate and starting point existing within each of the colored regions in Figure 5a. In addition, the remaining parameters for nondecision time and threshold were selected randomly from their corresponding estimated posterior distributions. Figure 5b and 5c shows the average predicted choice response time distributions under accuracy (5b) and speed (5c) emphasis instructions, respectively. These figures show that the high drift rate region produces responses that are both more accurate and faster than the low drift rate region. In fact, for the low drift rate region, the model predicts that responses will have accuracy that is near chance.

BOLD activation patterns as a function of drift rate. Figure 6 shows the mean predicted BOLD signal for each ROI during low (top row; the red region in Figure 5a) and high drift rate trials (bottom row; the green region in Figure 5a). Each ROI is represented as a "node" appearing on the nearest axial slice in Montreal Neurological Institute (MNI) coordinates and is labeled according to Table 1. The true shape and extent of each ROI is presented in the supplementary materials. The predicted BOLD signal is color coded according to the key on the right-hand side. Using Figure 6, we can identify brain regions that are associated with the slower information processing by locating ROIs that have high BOLD activity in the top row and low activity in the bottom row.

[7] The data are sparse relative to the number of model parameters, which would be problematic for the subsequent analyses. Hence, we relied on the PPD rather than the single-trial estimates themselves.

[8] In other words, the response time distributions were identical except that "correct" distributions resembled the "incorrect" distributions and vice versa.

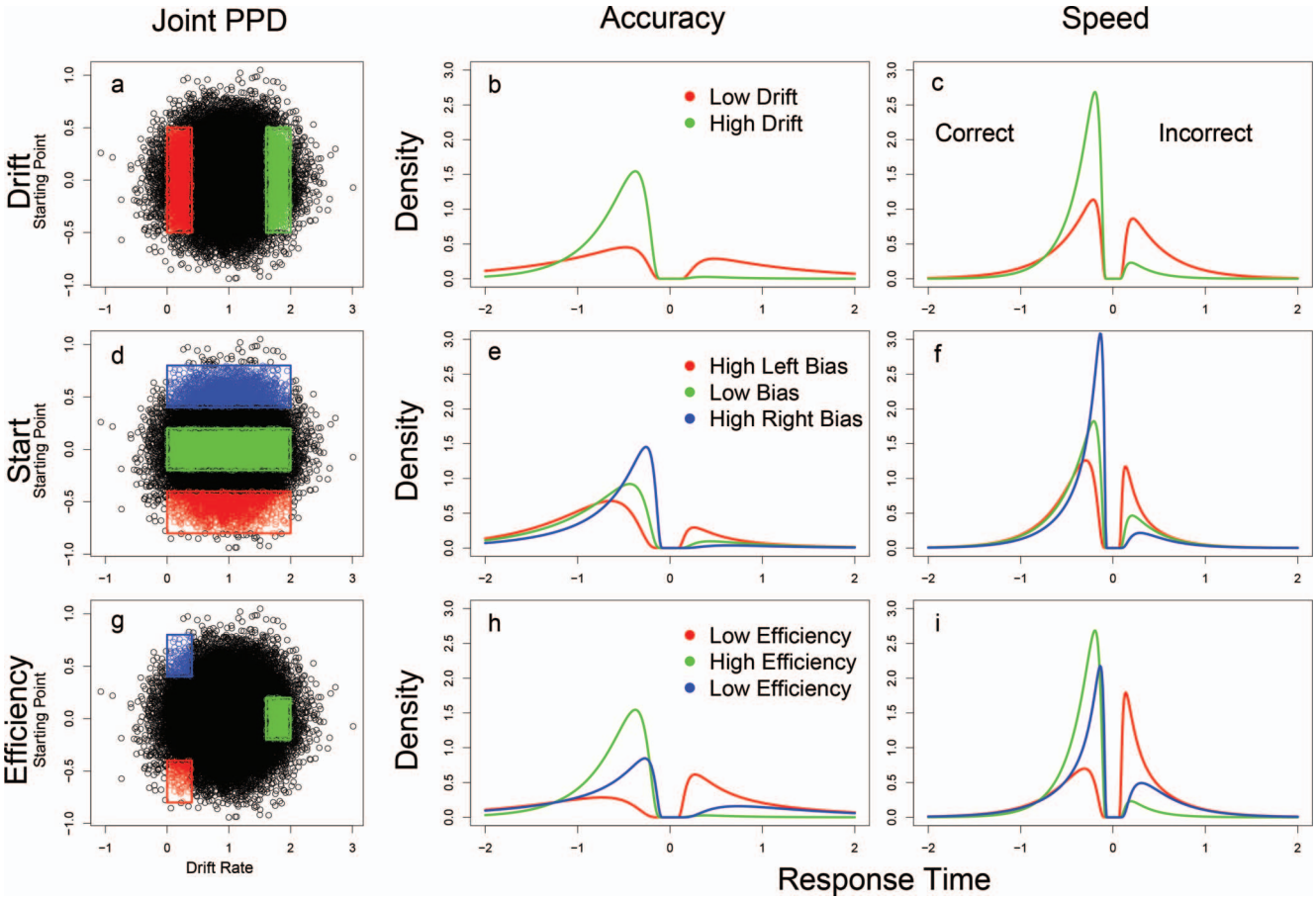

Figure 5. Model predictions for the behavioral data. Panels (a, d, g) show the joint posterior distributions of drift rate and starting point. Within each panel, a set of regions is defined to allow for the analysis of three cognitive constructs: (a) drift rate, (d) starting point (i.e., bias), and (g) efficiency. Panels (b, e, h) show the choice response time distributions for correct (left) and incorrect (right) responses under accuracy emphasis instructions, whereas Panels (c, f, i) show choice response time distributions under speed emphasis instructions. Panels (a, b, c) show the high and low drift rate regions, which are represented as the green and red regions/lines, respectively. Panels (d, e, f) show high- and low-bias regions, which are represented as blue (rightward response bias), red (leftward response bias), and green (low-bias) regions/lines, respectively. Panels (g, h, i) show the high- and low-efficiency regions, which are represented as green, and blue (rightward response bias) and red (leftward response bias) regions/lines, respectively. Note that only model predictions for rightward stimuli are shown and that the starting point is shown on the logit scale. PPD = posterior predictive distribution.

Default mode network. An appealing interpretation of these findings lies in the idea of the default mode network (DMN). Consistent patterns have emerged from studying the relationship between prestimulus brain activity and subsequent stimulus information processing. One pattern of activity that has a reliable effect on decision making is the DMN. The DMN is a network of brain regions that are active during off-task behaviors such as mind wandering, self-referential thought, or task-independent introspection (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Eichele et al., 2008; Gusnard & Raichle, 2001; Raichle et al., 2001; Raichle & Snyder, 2007; Weissman, Roberts, Visscher, & Woldorff, 2006). The DMN was originally observed as a pattern of brain activity present during the resting states of experiments (Raichle et al., 2001). Since this initial observation, the very notion of a DMN has sparked substantial controversy (Fair et al., 2008; Pollack & Devlin, 2007). Our view is that the existence of a DMN has drastic implications for how observers perform various tasks. For example, when the DMN is active, observers may engage in task-unrelated behaviors, creating suboptimal integration of stimulus information. Activation of the DMN might manifest behaviorally as an increase in errors (Eichele et al., 2008) or a slowing of the response times (Weissman et al., 2006). As such, the DMN has stimulated a significant amount of research exploring how activation in its subcomponents might relate to behavioral performance.

The general approach to studying the DMN is to relate (e.g., correlate) behavioral measures such as response accuracy, response time, or self-reports of awareness to brain activation patterns, such as the BOLD response. As examples, Weissman et al. (2006) used a local/global letter identification task to examine how

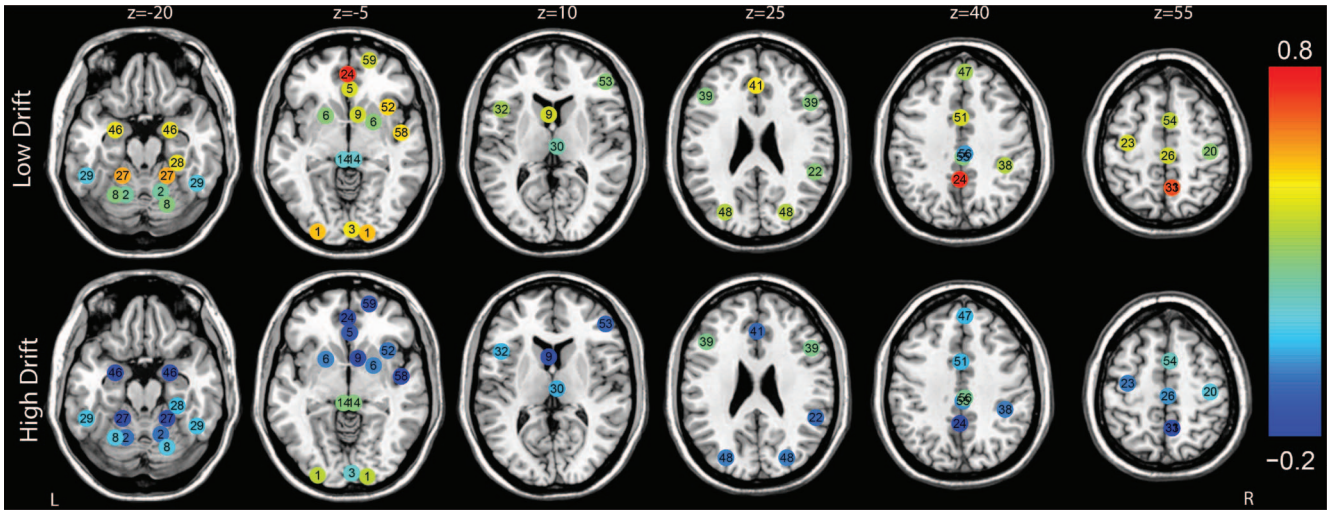

Figure 6. Brain activity as a function of the predicted rate of information processing on the single-trial level. The columns correspond to six axial slices, moving from ventral to dorsal surfaces. Each node corresponds to a region of interest, and the node's color represents the degree of activation for slow (i.e., low drift rates; top row) and fast (i.e., high drift rates; bottom row) information processing. In our model, drift rate serves as a proxy for the rate of stimulus information processing. High and low drift rate regions were defined by their relative locations in the marginal posterior distribution of the drift rate parameter (i.e., start point variability was integrated out; see Figure 5a).

response times related to brain activity. Their primary result was the identification of several regions that, when activated, produced longer response times (i.e., BOLD activity positively correlated with response time): anterior cingulate cortex, right middle frontal gyrus, and right inferior frontal gyrus. Eichele et al. (2008) found that patterns of reduced deactivation across trials in a region consisting of the inferior precuneus, posterior cingulate cortex, and retrosplenial cortex tended to predict future errors. In addition, they found decreased activity preceding errors in the posterior middle frontal cortex, orbital gyrus, inferior frontal gyrus, and SMA. Finally, Christoff et al. (2009) found that mind wandering as measured by self-reports and task accuracy was associated with activation of the dorsal and ventral anterior cingulate cortex, precuneus, temporoparietal junction, dorsal rostromedial and right rostrolateral prefrontal cortex, posterior and anterior insula, and bilateral temporopolar cortex, regions that are typically associated with the DMN.

Although the aforementioned studies have been instrumental in identifying the subcomponents of the DMN, they still leave much to be desired. First, the majority of previous examinations have explored how only a single behavioral measure relates to the DMN. Basing an inference on a single behavioral measure provides little constraint on the number of possible alternative explanations. For example, comparing neural activity during correct and incorrect trials only tells us *which* brain regions are (highly) active for each accuracy outcome. Such a comparison does not differentiate between, say, fast and slow errors, a feature of the data that has played an important role in differentiating competing psychological theories (Donkin, Nosofsky, Gold, & Shiffrin, 2013; Province & Rouder, 2012; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999). Second, previous examinations have employed discriminative analytic techniques, such as regression or Granger causality analysis. Discriminative approaches make no direct connection to an explicit cognitive theory because they are designed for data-driven analytic procedures (Bishop & Lasserre, 2007). As a result, discriminative models cannot enhance our understanding of *how* brain regions affect the decision process or *why* an active brain region is harmful to task performance mechanistically. On the other hand, generative models make explicit assumptions about the mechanisms underlying a cognitive process, and in so doing, they provide explanations for how neural activity relates to behavioral measures.

The DMN is generally thought of as a collection of brain regions that, when active, contribute negatively to overall task performance. However, Fox et al. (2005) observed patterns of functional connectivity at rest that suggest the brain is organized into two functional networks that are anticorrelated. The first network, which we call the "task-positive" network, consists of brain regions that, when active, contribute positively to overall task performance. The second network, which we call the "task-negative" network, works in an opposite manner and is consistent with general definitions of the DMN (Raichle & Snyder, 2007). To identify these networks, we must first separate task-negative behavior from task-positive behavior. In our model, we assume that task-negative behavior is most directly related to low drift rates, whereas task-positive behavior is related to high drift rates. We make this assumption because the drift rate parameter controls the rate of evidence accumulation, such that a high drift rate generally produces faster, more accurate responses, and a low drift rate produces slower, less accurate responses. Given the effect that the drift rate parameter has on the behavioral variables (see also Figure 5), we feel that it most closely maps onto the notion of the DMN.

To identify task-negative and task-positive networks, we need only determine the brain regions that have large deactivations during high drift rate trials. One way to assess the degree of deactivation is to simply take the difference between the predicted

Table 1
ROI Locations, Descriptions, and Mean Predicted Blood Oxygen Level-Dependent Level

ROI	Description	MNI coordinates *x*, *y*, *z*	LD	HD	LB	DB
1	Calcarine	−30, −95, 0; 18, −96, −5	0.580	0.384	0.470	0.049
2	Precentral gyrus	−17, −68, −21; 17, −65, −21	0.184	−0.074	0.078	0.300
3	Calcarine	3, −93, −9	0.474	0.147	0.317	0.005
5	vmOFC	1, 39, −12	0.418	−0.105	0.162	0.132
6	Putamen	−22, 14, −3; 24, 8, −3	0.276	0.009	0.146	0.050
8	Cerebellum	−27, −68, −22; 23, −77, −22	0.273	0.096	0.193	0.054
9	Caudate	−8, 15, 4; 9, 15, 2	0.428	−0.157	0.139	0.030
14	Thalamus/dorsal striatum	−5, −27, −3; 5, −27, −4	0.111	0.275	0.207	0.023
20	Postcentral gyrus	41, −25, 54	0.296	0.117	0.202	0.037
22	Medial temporal gyrus	59, −45, 19	0.280	−0.055	0.121	0.062
23	Pre/postcentral gyrus	−39, −17, 62	0.437	0.023	0.217	0.154
24	vmOFC/precuneus	−1, 53, −3; −1, −59, 34	0.745	−0.415	0.185	0.036
26	Paracentral gyrus	0, −29, 67	0.416	0.024	0.225	0.067
27	Cerebellum	−20, −50, −15; 23, −50, −18	0.607	−0.272	0.192	0.136
28	Fusiform gyrus	33, −38, −16	0.446	0.072	0.264	0.006
29	Inferior temporal gyrus	−55, −50, −17; 52, −57, −17	0.089	0.073	0.076	0.098
30	Thalamus	0, −15, 9	0.216	0.044	0.128	0.009
31	Precuneus	4, −59, 59	0.300	0.041	0.182	0.066
32	IFG pars triangularis	−54, 20, 10	0.349	−0.003	0.182	0.106
33	Precuneus	4, −59, 59	0.741	−0.318	0.217	0.100
38	Posterior IPS	43, −46, 46	0.357	−0.030	0.175	0.076
39	Middle frontal gyrus	−47, 26, 24; 55, 20, 25	0.274	0.263	0.261	0.207
41	Cingulate gyrus	2, 36, 20	0.493	−0.040	0.228	0.006
46	Parahippocampus	−27, −7, −26; 26, −7, −26	0.460	−0.180	0.147	0.022
47	Superior frontomedian cortex	4, 42, 39	0.358	0.028	0.193	0.015
48	Mid-occipital gyrus	−28, −83, 23; 31, −83, 25	0.368	−0.014	0.177	0.101
51	Cingulate sulcus	0, −1, 47	0.434	0.047	0.238	0.029
52	Anterior insula	38, 22, −10	0.535	−0.021	0.259	0.025
53	Frontopolar	47, 46, 4	0.294	−0.075	0.126	0.202
54	Pre-SMA/SMA	2, 4, 69	0.399	0.186	0.285	0.139
55	Precuneus	2, −38, 35	0.287	−0.004	0.145	0.075
56	Precuneus	4, −35, 45	0.007	0.263	0.131	0.108
58	Superior temporal gyrus	51, −2, −3	0.475	−0.190	0.153	0.025
59	Frontopolar	40, 48, −2	0.409	−0.131	0.147	0.215

Note. ROI = region of interest; MNI = Montreal Neurological Institute; LD = low drift; HD = high drift; LB = low bias; DB = average difference in activation across the two high-bias regions; vmOFC = ventromedial orbitofrontal cortex; IFG = inferior frontal gyrus; IPS = intraparietal sulcus; SMA = supplementary motor cortex.

activity levels during low and high drift rate trials. For example, in Figure 6, we can simply compare the BOLD activity in the top and bottom rows. When BOLD activity increases with increases in drift rate, the pattern of activity is consistent with the task-positive network. By contrast, when BOLD activity is high during low drift rate trials and low in high drift rate trials, the pattern of activity is consistent with the task-negative network. However, evaluating the magnitude of the deactivation does not incorporate the uncertainty about the predicted activation or the variability in the BOLD signal present for each individual ROI. One way to incorporate these uncertainties is through an examination of the PPD. For each ROI, the PPD contains a full distribution of the predicted BOLD activity for both high and low drift rate trials. Thus, we can estimate the probability that the BOLD activity during low drift rate trials is greater than the BOLD activity during high drift rate trials through numerical integration. Figure 7 shows these probabilities for each ROI, sorted in descending order. Higher probabilities in Figure 7 (i.e., the red region) are associated with task-negative behaviors, whereas lower probabilities (i.e., the green region) are associated with task-positive behaviors. ROIs located in the brownish-yellow area do not identify either behavioral characteristic with certainty. Brain regions most highly associated with the task-negative network are the anterior insula (52), cingulate gyrus (41), parahippocampus (46), caudate (9), ventromedial orbitofrontal cortex (vmOFC; 5), fusiform gyrus (28), vmOFC/precuneus (24), posterior intraparietal sulcus (38), and the superior temporal gyrus (58). Brain regions whose BOLD activity during high drift rate trials is greater than their activity during low drift trials are consistent with a task-positive network. We observed only two ROIs meeting this criterion: the precuneus (56) and the thalamus/dorsal striatum (14).

Although our particular ROI designation is unique, we can contrast our findings with prior research. The generally accepted set of subcomponents of the DMN consists of the following brain regions: medial temporal lobe, medial prefrontal cortex, anterior cingulate cortex, and the ventral precuneus. In our ROI set, the medial temporal lobe contains the parahippocampus (9) and the fusiform gyrus (28), whereas the medial prefrontal cortex contains the vmOFC (5) and the vmOFC/precuneus (24). In addition, there are several other ROIs in the nearby dorsal medial prefrontal cortex—such as the superior frontomedian cortex (47) and the cingulate gyrus (41)—and the orbitofrontal cortex, including a frontopolar region (59). For the anterior cingulate cortex, the cingulate gyrus (41) and the cingulate sulcus (51) are nearby (van Maanen et al., 2011). Finally, our ROI set

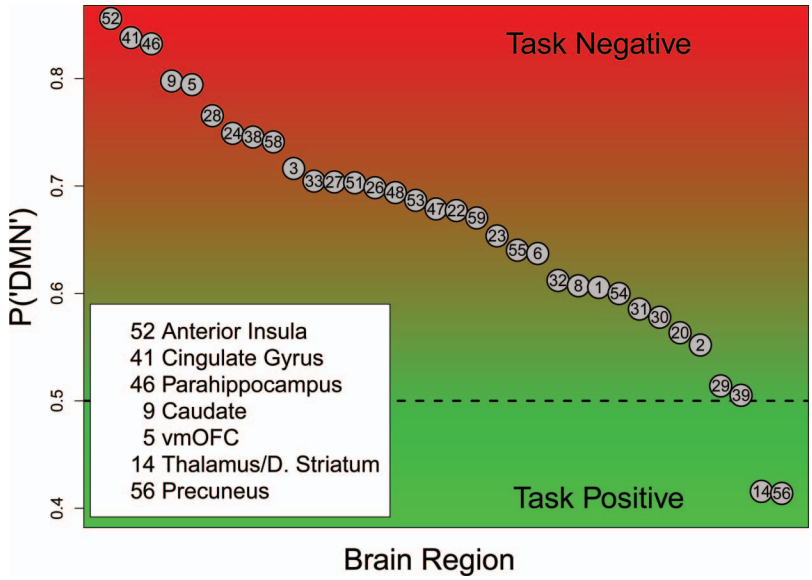

Figure 7. Probability of default mode network membership for each region of interest. For each region of interest, we evaluated the predicted probability that the blood oxygen level-dependent (BOLD) signal present for low drift rate trials would exceed the BOLD signal present during high drift rate trials. Higher probabilities indicate higher likelihood of task-negative network membership, whereas smaller probabilities indicate higher likelihood of task-positive network membership. DMN = default mode network; vmOFC = ventromedial orbitofrontal cortex.

contains five ROIs in the precuneus area: 31, 33, 55, 56, and partially the vmOFC/precuneus (24). Of these ROIs, ROI 31 and ROI 33 are located ventrally, whereas ROI 55 and ROI 56 are located anteriorly to the precuneus. Interestingly, while the BOLD signal present in ROI 55 showed no definitive behavioral influence, the BOLD signal present in ROI 56, which was located dorsally relative to ROI 55, exhibited task-positive behavior. Of these listed ROIs, all except for ROI 56 showed patterns of activity that were remarkably consistent with the DMN: high activity during low drift rate trials and low activity during high drift rate trials. However, when accounting for the variability in individual ROI activity (see Figure 7), we did not find strong evidence (i.e., as measured by the location of the posterior distribution) of DMN membership for the superior frontomedian cortex (47), the frontopolar region (59), or the precuneus (31, 33, 55). Although these results speak to the consistency of our model with prior research, they also indicate that our (Bayesian) analysis of the NDDM provides a slightly different interpretation once uncertainty and variability of BOLD activity are taken into account.

Starting Point Region Analysis

The second region analysis we performed was on the starting point. Figure 5d shows the joint posterior predictive distribution for single-trial drift rate and starting point. Three starting point regions are of interest. The first region corresponds to an area of low bias and is represented with the color green. To define the region, we selected starting point values that were within a small window (i.e., within 0.2 on the logit scale) of the point of unbiased responding at zero. Similar to the drift rate regions above, we selected a large range of drift rates so that the effects of drift on the decision could be fully integrated out. The resulting low-bias regions consisted of the starting points of $\boldsymbol{\omega} \in [0.45, 0.55]$, which is a proportion ranging from zero to 1. The second and third regions correspond to areas of high bias. The blue region represents a high bias for a rightward response (i.e., the correct response here), whereas the red region represents a high bias for a leftward response (i.e., the incorrect response). To define these regions, we selected areas that were equidistant from the point of unbiased responding and had the same width and range for drift rates as the low-bias region. The resulting high-bias regions consisted of the starting points of $\boldsymbol{\omega} \in [0.60, 0.69]$ for high rightward response bias and $\omega \in [0.31, 0.40]$ for high leftward response bias (on the probability scale).

As in the drift region analysis, we generated predictions for the behavioral data by randomly selecting values of drift rate and starting point that were contained within each of the regions in Figure 5d. Figure 5e and 5f shows the average predicted choice response time distributions under accuracy (5e) and speed (5f) emphasis instructions, respectively. The figures show that the most accurate responses are obtained in the high rightward bias region shown in blue. Because the contribution of the drift rate has been integrated out, the decision is being driven primarily by the starting point. Hence, when the starting point is nearer to the correct boundary (i.e., the "rightward" response boundary), the responses will tend to be more accurate than even the unbiased responding. However, if we were to plot the model predictions for leftward stimuli, the same blue region would produce the most inaccurate and slowest responses, relative to the other regions.

BOLD activation patterns as a function of starting point. The top row of Figure 8 shows the average BOLD signal for each ROI within the low-bias region (i.e., the green region in Figure 5d). Regions with the greatest activation (i.e., activations greater than 0.25) were the calcarine (1 and 3), pre-SMA/SMA (54), fusiform gyrus (28), middle frontal gyrus (39), and the anterior

insula cortex (52), whereas regions with the least activation (i.e., activation less than 0.14) were the inferior temporal gyrus (29), cerebellum (2), medial temporal gyrus (22), frontopolar cortex (53), thalamus (30), precuneus (56), and the caudate (9).

To better interpret the interaction of the BOLD signal and biased decision making, we used a measure of differential BOLD activity, given by

$$\boldsymbol{\delta}_k = \frac{1}{2}\sum_{i=1}^{2} \left| y_k^{LB} - y_{i,k}^{HB} \right|,$$

where Y_k^{LB} denotes the average BOLD signal for the kth ROI in the low-bias region, and $y_{i,k}^{HB}$ denotes the average BOLD signal for the kth ROI in the ith high-bias region. Thus, $\boldsymbol{\delta}$ is the average absolute difference in BOLD activity when moving from regions of low bias to regions of high bias for either alternative. The bottom row of Figure 8 shows $\boldsymbol{\delta}_k$ for each ROI, color coded according to the key on the right-hand side. Because a high value for δ implies either a high or low BOLD signal relative to the BOLD signal in the top row of Figure 8 (i.e., the BOLD signal during unbiased decisions), greater differential activation in the bottom row of Figure 8 is indicative that a given ROI contributes to starting point fluctuations that produce biased responding. Brain regions with the greatest average difference in activation (i.e., greater than 0.13) were the cerebellum (2), frontopolar cortex (59), middle frontal gyrus (39), frontopolar cortex (53), pre/postcentral gyrus (23), pre-SMA/SMA (54), cerebellum (27), and the vmOFC (5). For a given ROI, lower values in the bottom row of Figure 8 indicate that a particular ROI does not significantly affect the placement of the starting point. Regions with the lowest average difference in activation (i.e., less than 0.02) were the calcarine (3), fusiform gyrus (28), cingulate gyrus (41), thalamus (30), and superior frontomedian cortex (47).

One limitation of our analysis is that it does not disentangle the different types of high-bias response outcomes. Suppose, for example, a rightward motion stimulus is presented. If the model begins the trial with a high preference for a "rightward" response, it could produce either (a) a fast correct response or (b) a slow incorrect response. The latter of these two events could occur if the drift rate for the trial was relatively low, causing the response to be made on the basis of noisy stimulus integration. By examining the posterior distribution of a model that does not explicitly condition on accuracy of the response, we cannot explore the role of the BOLD activity on the accuracy of highly biased responses. Although this is a limitation of our analysis, it is not a limitation of the model. There are two ways we could examine the contribution of bias in greater detail. The first is a data-driven approach in which we would simply observe the pattern of BOLD activity present during trials where events (a) and (b) occurred. For these data, this approach is problematic because the limited number of observations would make the interpretation of such results difficult to substantiate. The second approach would be to reparameterize the model so that the thresholds in the model corresponded to the correct and incorrect response alternatives, rather than leftward and rightward response alternatives. Such a procedure would assume that rightward and leftward stimuli were equally preferable, and this assumption is reasonable for these data. However, the focus of the present study was to examine the degree to which brain regions were related to the rate of information processing—regardless of starting point—and so neither solution above was investigated in detail.

The drift rate and starting point regions analyses above are useful because they isolate parameters within the NDDM and map them directly to patterns of neural activity. However, one may also be interested in identifying how the brain executes processes that

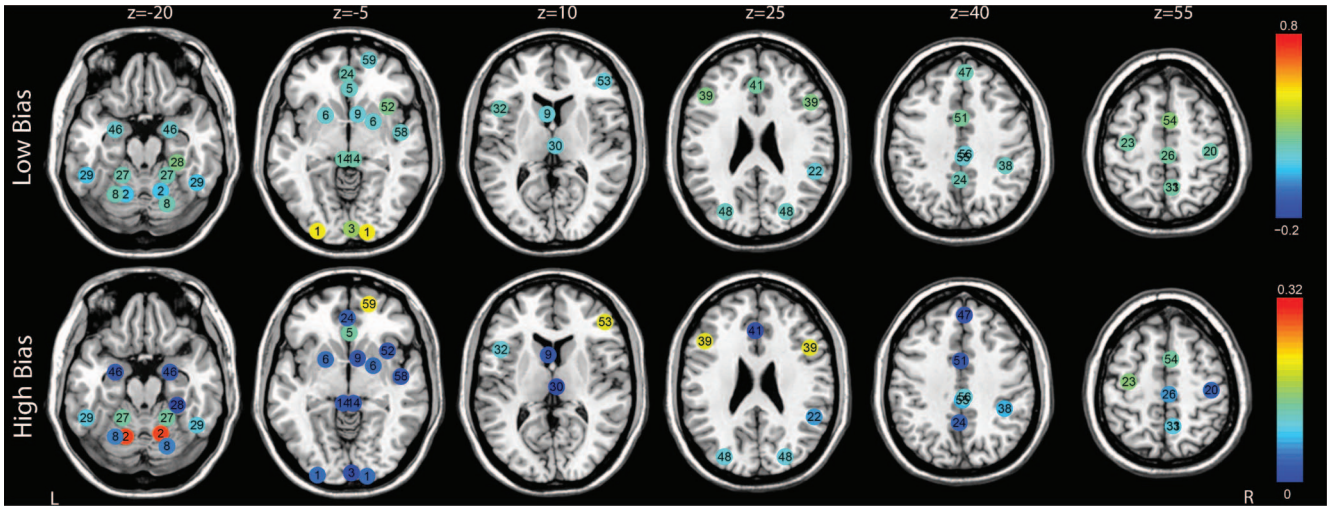

Figure 8. Brain activity patterns as a function of the degree of bias. The columns correspond to six axial slices, moving from ventral to dorsal surfaces. Each node corresponds to a region of interest, and the node's color represents the degree of activation. The top row shows the pattern of activity present during unbiased trials. The bottom row shows the average absolute difference in activation between trials with high and low bias. In our model, the initial starting points relative to $w = 0.5$ serve as a proxy for the degree of bias in the decision. High- and low-bias regions were defined by their relative locations in the marginal distribution of the starting point parameter (i.e., the drift rate variability was integrated out; see Figure 5d).

are more conceptual in nature, which can be defined by parameter regimes within the model. For example, one may be interested in identifying patterns of brain activity that produce behavioral responses that are accurate and relatively fast but are driven by the observer's ability to process stimulus information and not by "lucky guesses." In the next section, we perform an analysis on such a concept, which we refer to as *efficiency*.

Efficiency Region Analysis

We defined efficiency as the ability to make fast, yet unbiased, decisions. Within the NDDM, fast unbiased decisions are only made when (a) the drift rate is high and (b) the starting point is equally close to each of the response thresholds. To define the high-efficiency region, we selected a relatively large value for the drift rate ($\xi = 1.8$) and a small range of values for the starting point, because we wanted to limit the high-efficiency region to an unbiased parameter regime. In our model, unbiased responding is obtained when $w = 0.5$ or $\text{logit}(w) = \omega = 0$ as shown in Figure 5g. The high-efficiency region is represented in Figure 5g as the green area. To select low-efficiency regions, we first began by selecting relatively small values for the drift rate ($\xi = 0.2$). To define regions corresponding to biased decision making, we noted that bias in the decision is present if either alternative is preferred prior to stimulus onset. In the model, a preference for a "rightward" response is obtained by increasing the starting point w above 0.5. By contrast, preference for a "leftward" response is obtained by decreasing the starting point below 0.5. Thus, we defined two equally biased regions—one region corresponding to a high "rightward" response preference (i.e., the blue region in Figure 5g where $w = 0.65$) and one region corresponding to a high "leftward" response preference (i.e., the red region in Figure 5g where $w = 0.35$).

Figure 5h and 5i shows the average of these predicted distributions, where each distribution is color coded to coordinate with the regions defined in Figure 5g. Under both accuracy and speed emphasis instructions, we see that the model predicts correct responses with highest probability in the high-efficiency region (green). The second highest probability of a correct response is obtained in the blue low-efficiency region, and the lowest probability for a correct response is obtained in the red low-efficiency region. In fact, under both accuracy and speed emphasis instructions, the model predicts that incorrect responses are more likely than correct in the red low-efficiency region. The reason for this particular ranking of efficiency regions is because the blue low-efficiency region is nearer to the "rightward" response boundary, which is the correct response here.[9] For the red low-efficiency region, because the model begins near enough to the "leftward" response boundary and the drift rate is low enough, most of the responses spuriously reach the incorrect boundary.

Note that our particular definition of efficiency depends not only on response time but also on accuracy. In part, we formed our definition based on the interaction between drift rate and starting point and because a highly biased response does not imply inaccurate responding. For example, Figure 5h shows that the blue low-efficiency region actually produces faster correct responses than the high-efficiency region. However, the blue low-efficiency region also produces many more errors than the high-efficiency region, and the errors made by the blue low-efficiency region are slower than the errors made by the high-efficiency region. In addition, the drift rate interacts with the starting point such that a high enough drift rate can help prevent errors that would have resulted from a large initial bias.

Identifying efficient BOLD activation patterns. Figure 9 shows the average BOLD signal for each ROI within the high rightward bias low-efficiency region (i.e., the blue region in Figure 5g; top row), the high leftward bias low-efficiency region (i.e., the red region in Figure 5g; middle row), and the high-efficiency region (i.e., the green region in Figure 5g; bottom row). Each ROI is represented as a "node" appearing on the nearest axial slice in MNI coordinates and is labeled according to Table 1. Some ROIs containing more than one brain region are represented as multiple nodes. For example, calcarine (1) was defined bilaterally, and so two nodes are used to represent this ROI in the figure. The predicted BOLD signal is color coded according to the key on the right-hand side. High BOLD activity in the top and middle rows combined with low BOLD activity in the bottom row indicates that a particular ROI contributes negatively to efficient decision making. Figure 9 shows that greater efficiency can be achieved when most ROIs have the greatest deactivation (i.e., lowest activation), but the degree of deactivation varies across ROIs. For example, regions corresponding to the vmOFC (5, 24) require greater deactivation than do regions corresponding to either the inferior temporal gyrus (28) or lateral portions of the cerebellum (8). The greatest deactivations (i.e., having a deactivation greater than 0.6) were observed for the following regions (from greatest to least): precuneus (33), vmOFC/precuneus (24), anterior portions of the cerebellum (27), frontopolar region (59), vmOFC (5), parahippocampus (46), superior temporal gyrus (STG; 58), caudate (9), and a medial portion of the cerebellum (2). Furthermore, greater efficiency is obtained with greater activation of the thalamus and dorsal striatum (14) and a dorsal medial area of the precuneus (56). Some ROIs exhibited a nonmonotonic relationship with efficiency; for example, high activation or high deactivation of the middle frontal gyrus (39) and the inferior temporal gyrus (29) produced inefficient responses, whereas midrange activations (i.e., around 0.1) produced highly efficient responses.

To this point, we have shown that our (generative) modeling approach allows for the identification of brain regions whose activation correlates with parameters of a cognitive model. Another way we can evaluate the merits of NDDM relative to the DDM is by examining how the models perform on a cross-validation test. The objective here is to show that the NDDM can use the information provided by the neural data to generate better predictions for behavioral data than the DDM. Because we have already shown that the NDDM provides a better qualitative explanation of the data than the DDM, such a test—if successful—would demonstrate that the NDDM provides an enhanced view of the data that is not subject to poor generalization (Myung & Pitt, 2002).

Comparing the Models

In this section, we perform a predictive modeling test via cross-validation. First, we randomly removed 100 trials—referred to as the "test" data—from our data set and refit both models to the remaining "training" data. Once the (new) posteriors for both models had been

[9] Recall that only the predictions for rightward motion stimuli are shown in Figure 5.

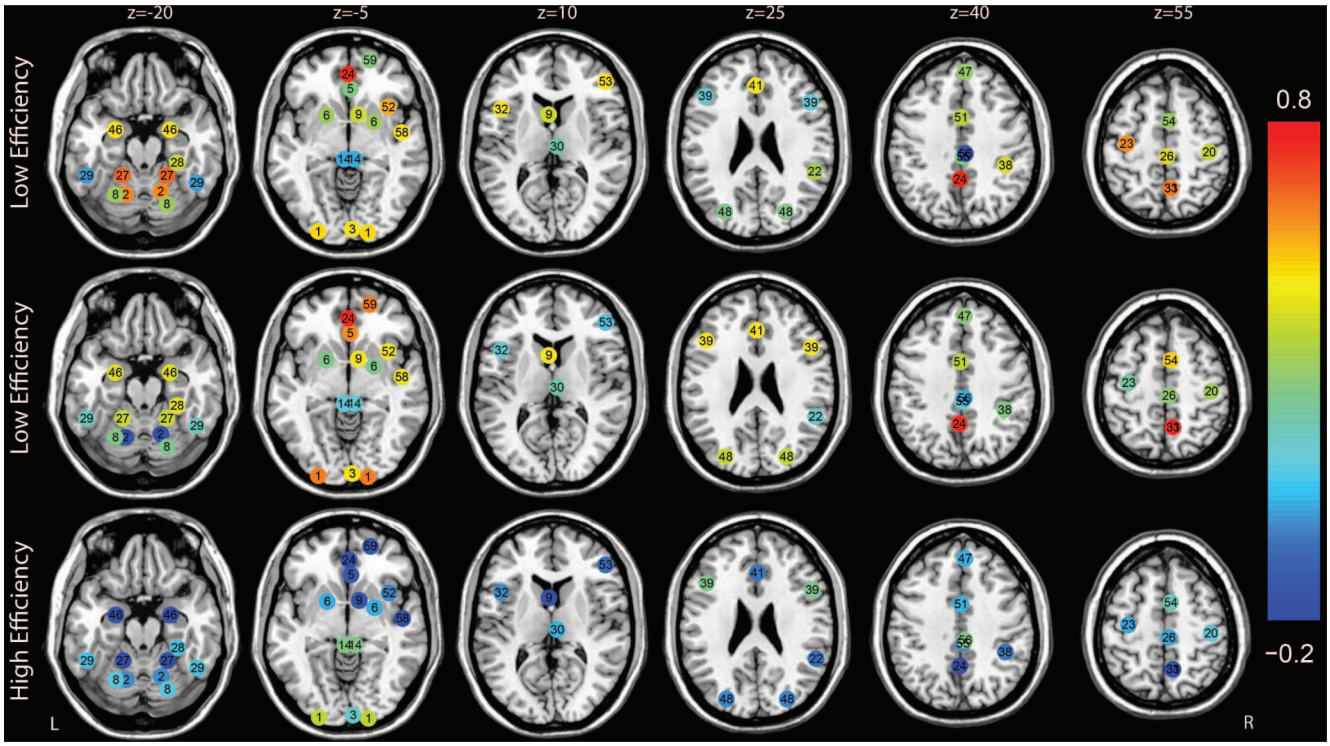

Figure 9. The efficiency of information processing as a function of blood oxygen level-dependent (BOLD) activation. Efficiency is defined as the ability to produce fast, yet unbiased, responses, which is determined by the location in the joint posterior distribution of drift rate and starting point (see Figure 5g). The columns correspond to six axial slices, moving from ventral to dorsal surfaces. Each region of interest is represented as a node in the figure color coded to reflect its degree of BOLD activation. The top row corresponds to the low-efficiency region with high rightward bias (i.e., the blue region in Figure 5g), the middle row corresponds to the low-efficiency region with high leftward bias (i.e., the red region in Figure 5g), and the bottom row corresponds to the high-efficiency region (i.e., the green region in Figure 5g).

estimated, we used the posterior estimates for the hyperparameters to generate PPDs for the single-trial drift rate and starting point parameters, consisting of 1,000 random samples. For the NDDM, this generation process conditioned on the information in the neural data for each test trial as described in Turner (2013). As an illustrative example, Figure 10a and 10b shows the PPDs for the (single-trial) drift rate parameter $\boldsymbol{\xi}$ and (single-trial) starting point parameter $\boldsymbol{\omega}$, respectively, for the NDDM (green histograms) and the DDM (red histograms) on Test Trial 82. When comparing the PPDs, we observe two things. First, the PPDs for both parameters in the NDDM are more constrained (i.e., have smaller variance) relative to the DDM model parameters, especially for the drift rate parameter $\boldsymbol{\xi}$ (see Figure 10a). For the NDDM, the means over 100 test trials of the standard deviations of the posterior distributions for $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ were 0.248 and 0.209, respectively, whereas for the DDM, the standard deviations were 0.407 and 0.244, respectively. Clearly, the posterior distributions of the model parameters for the NDDM have less variance relative to the DDM, where a larger discrepancy exists between drift rates. As we observed in the simulation study, the posteriors in the NDDM are more constrained because of the information provided by the neural data. Turner (2013) showed that the additional constraint on the behavioral model parameters is guaranteed as long as a nonzero correlation exists between *at least one* neural source and a particular latent variable. In our analyses, we observed several strong correlations between both model parameters and many neural sources (see Figure 2), and as a consequence, the information in the neural data better constrains the cognitive model. Second, we observe that the means of the PPDs for both model parameters differ between the NDDM and the DDM. The difference in means is also caused by the neural data. Similar to our observations in the simulation study above, for the NDDM, the means of the single-trial posterior distributions fluctuate according to the neural data, whereas for the DDM, the means fluctuate only slightly, because they are being generated only from the prior distribution (i.e., it ignores the neural data).

Once the PPDs had been generated for each model parameter, we used the 1,000 random samples to generate a PPD for the behavioral data. Figure 10c shows the PPDs in data space, where green corresponds to the NDDM and red corresponds to the DDM (i.e., as in Figure 10a and 10b). The 95% credible set for the response time distributions of correct (left) and incorrect (right) responses is shown for each model, along with the median of the PPDs shown in corresponding color. Comparing the two credible sets, we can see that the predictions under the NDDM are less variable relative to the DDM. The dashed vertical line represents the data for Test Trial 82, which was a correct response with a response time of 525 ms. To compare the accuracy of the predictions for both models, we can evaluate the

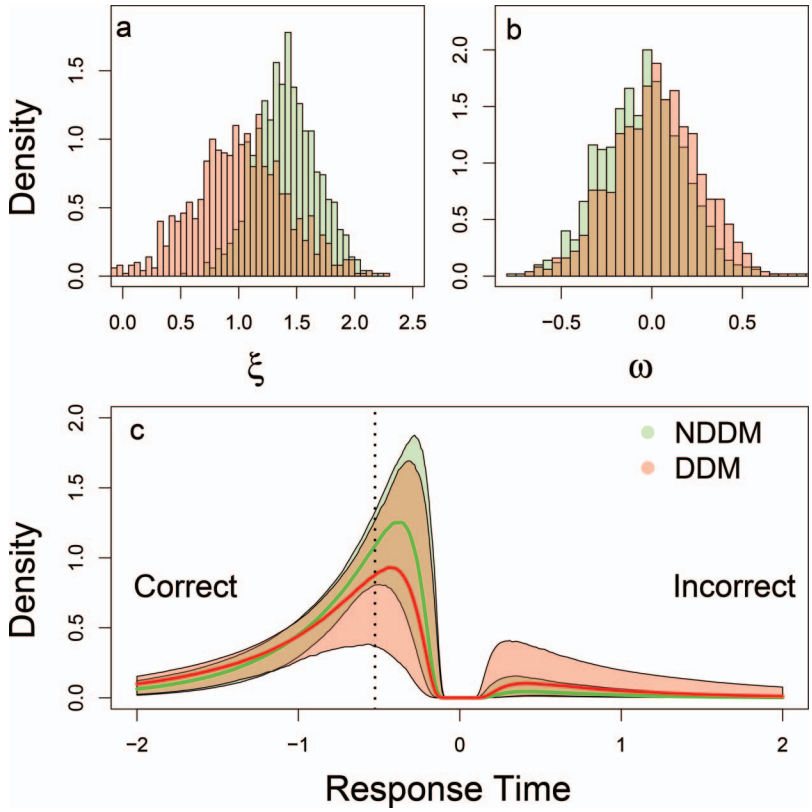

Figure 10. Posterior predictive distributions of model parameters and data variables for Test Trial 82. In all panels, neural drift diffusion model (NDDM) distributions are shown in green, whereas drift diffusion model (DDM) distributions are shown in red. The top panels show the posterior predictive distributions for (a) the (single-trial) drift rate ξ and (b) the (single-trial) starting point ω. Panel c shows the posterior predictive distribution for the choice response time distributions for correct (left) and incorrect (right) responses. For each model, 95% credible sets are plotted along with the best prediction illustrated with a solid line of corresponding color. The dashed vertical line represents the test trial data (i.e., a correct response with a response time of 525 ms).

density of the test trial data under the response time curves. If a model places higher density at the location of a particular test trial data point, the predictions are more accurate because the model believes the test data are most likely to occur at this location. Hence, the higher the density around the test trial data, the more accurate the predictions from the model. For Test Trial 82, the NDDM has a higher density than the DDM, making it the preferred model for this test trial.

We can repeat the model comparison process illustrated in Figure 10 for the remaining test trials. However, to do so, we will evaluate the likelihood of each test trial under the collection of probability density functions (i.e., the entire shaded region in Figure 10c). Evaluation of the likelihoods under all probability density functions in the PPDs produced a distribution of 1,000 likelihood values for each model. To compare the distributions between the two models, we used a Monte Carlo procedure for estimating the probability that the distribution of likelihood values for the NDDM was higher than the distribution of likelihood values for the DDM (see Robert & Casella, 2004).

Figure 11 shows the probability that the NDDM is the preferred model for each of the 100 test trials, ranked according to increased preference for the NDDM. Each probability is color coded to represent whether the test trial resulted in a correct (black dots) or incorrect (gray dots) response. The vertical arrow points to Test Trial 82, which was used as the illustrative example in Figure 10. A horizontal reference line is plotted to correspond to equal preference for each model at 0.5, and a vertical reference line is plotted to correspond to 50% of the test trials. The figure shows that 63 of the 100 trials are better predicted by the NDDM, which is indicated by observing the number of dots above the horizontal line. In our analyses, the NDDM performed better than the DDM on 83% of the incorrect trials but only 57% of the correct trials.[10]

Discussion

In this article, our goal was to develop a model that could use prestimulus measures of brain activity to better constrain and inform the mechanisms assumed by a cognitive model of choice

[10] In two replications of the cross-validation study, we noted that the NDDM performed consistently better on the incorrect trials than on the correct trials. We speculated that the performance difference is due to the NDDM's ability to make use of the task-negative network in predicting subsequent behavior. However, additional data will need to be collected to ensure that these initial findings are not spurious.

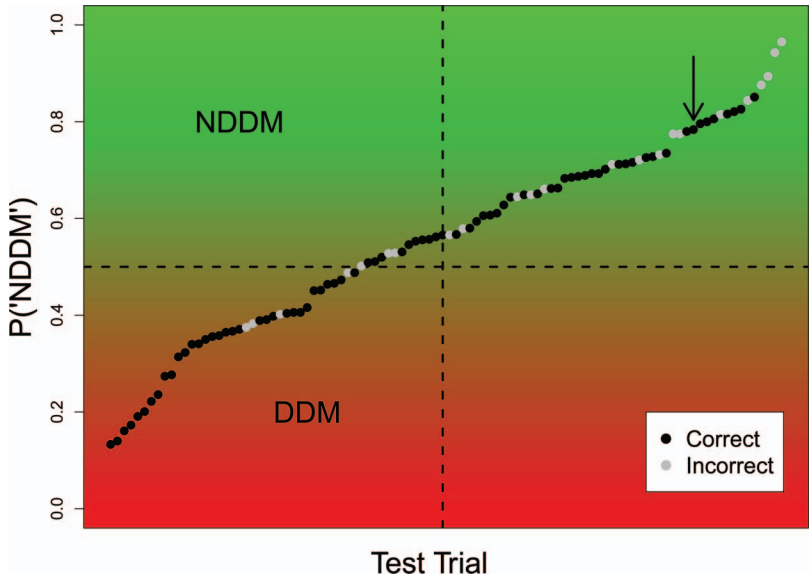

Figure 11. Cross-validation test comparing the neural drift diffusion model (NDDM) to the drift diffusion model (DDM). The numerical estimates for model predictability are plotted in ascending order of NDDM preference. Each point is color coded to indicate that the trial was either correct (black dots) or incorrect (gray dots). The horizontal reference line at 0.5 represents equal model preference, whereas the vertical reference line at 50 corresponds to half the data (i.e., 50 trials on either side). The arrow identifies the likelihood comparison made on Test Trial 82 (see Figure 10).

response time. To accomplish this, we developed the NDDM by extending the joint modeling framework to link neural and behavioral data at the single-trial level (Turner, Forstmann, et al., 2013). Our model captures many sources of variability simultaneously and takes full advantage of the set of available constraints. Structuring the model in this way creates an interesting dynamic between neural and behavioral measures. For example, our model predicts behavior not only on the basis of past behavioral data but also on the basis of neural data observed on a particular trial.

To test our model, we directly compared it to a version of the DDM (Ratcliff, 1978) in a simulation study. Comparing the DDM to the NDDM is an important test because the models are equivalent except that the DDM neglects the information in the neural data. Hence, if the NDDM could outperform the DDM, it would suggest that our model-based approach is an advancement to the traditional approach of cognitive modeling. We provided both models with a limited number of training examples and then tested their ability to predict future data (i.e., the "test" data). To compare the accuracy of the models, we evaluated the correlations of the true single-trial parameters with the maximum a posteriori predictions of the models (see Figure 3). We found that the predictions of the NDDM significantly correlated with both single-trial parameters, whereas the DDM's predictions were not significantly correlated with either single-trial parameter.

We then used our model to further examine how brain activity relates to behavioral data (and vice versa) in a perceptual decision-making task. We found that specific combinations of start point and drift rate parameters predicted different behavioral decision dynamics as well as neural activation in different ROIs. Specifically, we found that the combination of high drift rates and no bias (which we defined as *efficiency*) was related to fast responses with few errors. Additionally, high efficiency was related to deactivation of many ROIs. These effects were decomposed into an effect related to the start point parameters and an effect related to the drift rate parameters. Starting point fluctuations were related to biased behavior (i.e., fast correct and slow incorrect responses or vice versa), as well as with activation in the cerebellum, frontopolar cortex, middle frontal gyrus, pre/postcentral gyrus, pre-SMA, and the vmOFC. In contrast, higher drift rates predicted faster and more correct choices. This was associated with decreased activation in the precuneus. Conversely, low drift rates predicted slow and incorrect responses and increased activation in the precuneus and cerebellum.

We also investigated the role of the DMN on the basis of the NDDM. For our purposes, it was convenient to identify both task-negative and task-positive networks, whose components are anticorrelated (Fox et al., 2005). Activation of brain regions within the task-negative network tended to produce responses that were inaccurate. When the responses were accurate, they were generally slower relative to the full response time distribution. We then assessed the degree to which each ROI within the task-negative network was associated with the mechanisms in our model, specifically, the rate of information processing. We do not necessarily believe that the drift rate parameter corresponds exclusively to the DMN. It is entirely plausible that low drift rate trials contain a mixture of brain areas that produce task-negative behavior, such as the DMN *plus* brain areas where the influence of the stimulus is expressed. Regardless, the drift rate parameter does give us the best chance of identifying regions associated with the conventional notion of the DMN. We found that many of the components of the task-negative network were exclusively associated with predicting the rate of information processing—specifically, the precuneus,

vmOFC/precuneus, anterior cerebellum, parahippocampus, STG, caudate, and the anterior insula. By contrast, the pre-SMA/SMA, middle frontal gyrus, and the medial portion of the cerebellum mostly influenced the starting point. Our model could not cleanly differentiate which mechanism had the highest association with a frontopolar region (59). Surprisingly, the task-positive network included only two components—the thalamus/dorsal striatum and the medial precuneus. We found that whereas the thalamus/dorsal striatum (14) was exclusively related to drift rate, the medial precuneus (56) was associated with both drift rate and starting point.

There are also a few differences between our results and prior research. Jointly modeling the neural and behavioral data allowed us to determine whether trial-to-trial fluctuations in BOLD activity were more similar to trial-to-trial fluctuations in starting point or drift rate. We assumed that the notion of a DMN was better ascribed mechanistically to the drift rate parameter, enabling us to identify brain regions through parameters in the model. Thus, when a brain region is not highly associated with the drift rate, we argue that it functions differently from what is required by current definitions of the DMN. With this view in mind, the major differences between our results and prior research primarily fall onto four brain regions. First, the middle frontal gyrus (39) and the pre-SMA (54) were unrelated to the drift rate but were highly related to the starting point. Second, we found that the right inferior temporal gyrus was unrelated to drift rate, but there was some small evidence that it was related to starting point. Third, we found that the middle temporal gyrus (22) was related to drift when using simple differencing, but when overall uncertainty in the predicted BOLD activity was taken into account, we did not observe a significant mapping to the drift rate or starting point. Other reasons for these differences could be due to inconsistent ROI specifications across studies, the task used in the experiment, or the incorporation of multiple behavioral measures. Specifically, we are unaware of any study in neuroscience that combines choice and response time into a single analysis.

We can also compare our results to those of van Maanen et al. (2011), although the implementation of the single-trial LBA (STLBA) model and the NDDM differs strongly, and it would not be fair to expect an exact correspondence between the results. However, because conceptually the models are closely related, it is expected that both models make similar predictions for those ROIs in which the signal is strongest (i.e., the relation with the latent variable is strongest). This is indeed what we found. van Maanen et al. report a correlation between single-trial starting point and the middle frontal gyrus, pre-SMA/dorsal anterior cingulate cortex, putamen, anterior cingulate cortex, superior occipital gyrus, and precuneus. The parameter in NDDM that is mostly related to a single-trial start point is the bias parameter, because both parameters capture the distance to threshold on a trial-by-trial basis. Indeed, the NDDM reveals that SMA/pre-SMA and middle frontal gyrus are associated with a large difference in bias, which can be interpreted as those trials in which the initial activation is already close to threshold.

Although there are some similarities, the NDDM and the STLBA model also report some differences. These nonoverlapping regions may be due to differences in the functional role of these areas, as represented by the slightly different parameters of both models. For example, (van Maanen et al., 2011) found that the single-trial starting point correlates with single-trial BOLD in the putamen, but this was not found with the NDDM. One possible interpretation is that the STLBA starting point parameter indexes the overall threshold, and the NDDM starting point parameter indexes the distance to threshold of one accumulator. This difference would lead to the interpretation that the putamen is involved in setting overall response caution, a notion generally supported by the literature (e.g., Forstmann et al., 2008, 2010). In a similar way, the implication of the vmOFC in corresponding to bias by the NDDM (but not the STLBA model) may be in support of the bias-related function of this brain area (Mulder et al., 2013).

Although we have shown that the NDDM provides a powerful explanation of how neural sources are related to mechanisms within the model, it is also important to show that the NDDM is advantageous in predictive modeling. One reason for this is the counterbalance between model complexity, goodness of fit, and generalizability (e.g., Myung & Pitt, 2002). Without explicitly showing the NDDM's generalizability, one may wonder if the model is simply more complex than the DDM. To demonstrate the NDDM's potential in predictive modeling, we compared the NDDM to the DDM in a cross-validation test. We first removed 100 trials (i.e., the "test data") at random from the experimental data. We then refit the models to the remaining "training data" and used the estimated posterior distributions to generate predictions for the test data. Figure 10 shows how the information in the neural data constrains the model and limits its flexibility in generating predictions. As a consequence of this additional constraint, the NDDM outperformed the DDM in the cross-validation test.

Alternative Model-Based Approaches

Although our model-based approach is unique in application, it is not unique in motivation. Many authors have advocated for the power of reciprocal relations between neuroscience and mathematical modeling (cf. Forstmann, Wagenmakers, et al., 2011). Currently, a variety of methodologies and theoretical frameworks exist for accomplishing this reciprocity. Some of these approaches are theoretical in nature, where the physiology of neural function alone inspires the development of cognitive architectures (e.g., Mazurek et al., 2003McClelland & Rumelhart, 1986; O'Reilly, 2001, 2006; O'Reilly & Munakata, 2000; Shadlen & Newsome, 2001; Usher & McClelland, 2001). Other approaches aim to incorporate mechanisms that describe the production of neural data on top of an underlying cognitive model (e.g., Anderson et al., 2010; Anderson et al., 2008; Anderson et al., 2012; Anderson, Qin, Jung, & Carter, 2007; Fincham, Anderson, Betts, & Ferris, 2010) or use the raw neural data to directly replace certain mechanisms within the cognitive model (e.g., Purcell et al., 2010; Purcell, Schall, Logan, & Palmeri, 2012; Zandbelt, Purcell, Palmeri, Logan, & Schall, 2014).

As we discussed earlier, a particularly relevant approach to model-based neuroscience is the two-stage correlation approach (e.g., Cavanagh et al., 2011; Forstmann et al., 2008; Forstmann et al., 2010; Forstmann, Tittgemeyer, et al., 2011; Ho et al., 2012; Ho et al., 2009; Philiastides et al., 2006; Ratcliff et al., 2009; van Maanen et al., 2011; Wiecki, Sofer, & Frank, 2013). In this approach, behavioral model parameters are estimated and then correlated with a neural signature of interest. Cavanagh et al. used a two-stage approach in which the parameters of a hierarchical

drift diffusion model (HDDM; Wiecki et al., 2013) were regressed against important aspects of their EEG data. Cavanagh et al.'s approach is similar in spirit to the NDDM presented here. In fact, one can show that the joint modeling framework *subsumes* the two-stage correlation procedure, and by extension, the NDDM subsumes the regression approach of Cavanagh et al. This regression approach, although informative, still suffers from the problem of neglecting parameter constraint and ignores important relationships that might exist between neural sources, such as spatial proximity or neuroanatomy (i.e., a problem of multicollinearity). The NDDM, on the other hand, makes the interrelationships between sources an explicit part of the model that is used to amplify its predictive power.

Implementing the NDDM

Given the complexity of the NDDM, one may wonder how feasibly the NDDM can be implemented. We combine several methodological advancements to fit the model to data. First, we used the algorithm of Navarro and Fuss (2009) for efficient calculation of the model's first passage time distributions. Second, we use an efficient, scalable algorithm for performing Bayesian sampling of the joint posterior, called differential evolution with Markov chain Monte Carlo (ter Braak, 2006; Turner & Sederberg, 2012; Turner, Sederberg, et al., 2013). Third, to accomplish the hierarchical modeling, we first pick convenient transformations of the single-trial parameters so that they both have continuous, infinite support on the parameter space. For the drift rates, no transformation was necessary, whereas for the starting points, we transformed the relative starting point parameter $\boldsymbol{w}$ through a logistic function to produce $\boldsymbol{\omega}$. Once the parameters were on the appropriate space, we could assume that the parameters fluctuated from trial to trial according to a normal distribution. Assuming normality then made it possible to specify conditionally conjugate prior distributions for the hyperparameters (see Gelman et al., 2004, for similar derivations). Establishing conditionally conjugate relationships among the hyperparameters allowed us to reduce the dimensionality of the estimation problem by generating better proposals for the parameters in our fitting routine (cf. Robert & Casella, 2004).

Conclusions

In this article, we developed the NDDM as a way to *simultaneously* understand neural and behavioral data at the single-trial level. Using our model, we were able to provide a mechanistic interpretation of fMRI data through the lens of a cognitive model. We showed how the NDDM used the information in the neural data to generate significantly better predictions for behavioral data in a cross-validation test. Such a modeling framework signals an important advance for assessing and monitoring human performance in a variety of tasks, especially day-to-day operations with high attentional or "on-task" demands (e.g., flying a plane, driving a car). Our research efforts further highlight the utility of neuroimaging in cognitive modeling, and more specifically, they signal the importance of integrating neural and behavioral measures at the single-trial level.

References

Anderson, J. R., Betts, S., Ferris, J. L., & Fincham, J. M. (2010). Neural imaging to track mental states. *Proceedings of the National Academy of Sciences, 107,* 7018–7023.

Anderson, J. R., Carter, C. S., Fincham, J. M., Qin, Y., Ravizza, S. M., & Rosenberg-Lee, M. (2008). Using fMRI to test models of complex cognition. *Cognitive Science, 32,* 1323–1348.

Anderson, J. R., Fincham, J. M., Schneider, D. W., & Yang, J. (2012). Using brain imaging to track problem solving in a complex state space. *NeuroImage, 60,* 633–643.

Anderson, J. R., Qin, Y., Jung, K. J., & Carter, C. S. (2007). Information-processing modules and their relative modality specificity. *Cognitive Psychology, 54,* 185–217.

Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7,* 1129–1159.

Bishop, C. M., & Lasserre, J. (2007). Generative or discriminative? Getting the best of both worlds. *Bayesian Statistics, 8,* 3–24.

Boehm, U., Van Maanen, L., Forstmann, B., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage, 96,* 95–105.

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

Calhoun, V., Adali, T., Pearlson, G., & Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping, 14,* 140–151.

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience, 14,* 1462–1467.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106,* 8719–8724.

Craigmile, P., Peruggia, M., & Zandt, T. V. (2010). Hierarchical Bayes models for response time data. *Psychometrika, 75,* 613–632.

Criss, A. J., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64,* 316–326.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *Journal of Neuroscience, 31,* 1780–1789.

DeCarlo, L. T. (2011). Signal detection theory with item effects. *Journal of Mathematical Psychology, 55,* 229–239.

de Lange, F. P., Jensen, O., & Dehaene, S. (2010). Accumulation of evidence during sequential decision making: The importance of top-down factors. *Journal of Neuroscience, 30,* 731–738.

de Lange, F. P., van Gaal, S., Lamme, V. A., & Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision making. *PLoS Biology, 9,* e1001203.

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review, 120,* 873–902.

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., . . . Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences, 16,* 6173–6178.

Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., . . . Schlaggar, B. L. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences, 105,* 4028–4032.

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). New York, NY: John Wiley.

Fincham, J. M., Anderson, J. R., Betts, S. A., & Ferris, J. L. (2010). Using neural imaging and cognitive modeling to infer mental states while using and intelligent tutoring system. In *Proceedings of the Third International Conference on Educational Data Mining (EDM2010)*. Pittsburgh, PA.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., . . . Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences, 107,* 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,* 17538–17542.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience, 31,* 17242–17249.

Forstmann, B. U., & Wagenmakers, E. J. (2014). *An introduction to model-based cognitive neuroscience*. New York, NY: Springer.

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences, 15,* 272–279.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102,* 9673–9678.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience, 2,* 685–694.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274,* 427–430.

Heathcote, A., Brown, S. D., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review, 7,* 185–207.

Himberg, J., Hyvarinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage, 22,* 1214–1222.

Ho, T., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E. J., & Serences, J. T. (2012). The optimality of sensory processing during the speed-accuracy tradeoff. *Journal of Neuroscience, 32,* 7992–8003.

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

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review, 121,* 1–32.

Kac, M. (1962). A note on learning signal detection. *IRE Transactions on Information Theory, 8,* 126–128.

Kac, M. (1969). Some mathematical models in science. *Science, 166,* 695–699.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience, 28,* 3017–3029.

Kiviniemi, V., Starck, T., Remes, J., Long, X., Nikkinen, J., Haapea, M., . . . Tervonen, O. (2009). Functional segmentation of the brain cortex using high model order group PICA. *Human Brain Mapping, 30,* 3865–3886.

Liu, T., & Pleskac, T. J. (2011). Neural correlates of evidence accumulation in a perceptual decision task. *Journal of Neurophysiology, 106,* 2383–2398.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology, 23,* 2023–2027.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13,* 1257–1269.

McClelland, J., & Rumelhart, D. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.

Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology, 52,* 376–388.

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E. J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, Perception and Psychophysics, 75,* 1048–1058.

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences: A model-based review. *Neuroscience, 277,* 872–884.

Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience, 32,* 2335–2343.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage, 59,* 2636–2643.

Myung, I. J., & Pitt, M. (2002). Mathematical modeling. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 429–460). New York, NY: John Wiley.

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology, 53,* 222–230.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15,* 1729–1737.

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-Based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Science, 1104,* 35–53.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation, 13,* 1199–1242.

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

O'Reilly, R. C., & Munakata, Y. (Eds.). (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Peruggia, M., Van Zandt, T., & Chen, M. (2002). Was it a car or a cat I saw? An analysis of response times for word recognition. *Case Studies in Bayesian Statistics, 6,* 319–334.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficult and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience, 26,* 8965–8975.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News, 6,* 7–11. Retrieved from http://CRAN.R-project.org/doc/Rnews/

Pollack, R., & Devlin, J. T. (2007). On the fundamental role of anatomy in functional imaging: Reply to commentaries on 'In praise of tedious anatomy'. *NeuroImage, 37,* 1073–1082.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156.

Pratte, M. S., & Rouder, J. N. (2011). Hierarchical single- and dual-process models of recognition memory. *Journal of Mathematical Psychology, 55,* 36–46.

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences, 109,* 14357–14362.

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.

Purcell, B., Schall, J., Logan, G., & Palmeri, T. (2012). Gated stochastic accumulator model of visual search decisions in FEF. *Journal of Neuroscience, 32,* 3433–3446.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98,* 676–682.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage, 37,* 1083–1090.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences, 106,* 6539–6544.

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

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Comparing connectionist and diffusion models of reaction time. *Psychological Review, 106,* 261–300.

Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods*. New York, NY: Springer.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912.

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.

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.

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage, 77,* 157–165.

Tosoni, A., Galati, G., Romani, G. L., & Corbetta, M. (2008). Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. *Nature Neuroscience, 11,* 1446–1453.

Treisman, M., & Williams, T. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

Turner, B. M. (in press). Constraining cognitive abstractions through Bayesian modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience*. New York, NY: Springer.

Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage, 72,* 193–206.

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., Brown, S., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods, 18,* 368–384.

Turner, B. M., Van Zandt, T., & Brown, S. D. (2011). A dynamic, stimulus-driven model of signal detection. *Psychological Review, 118,* 583–613.

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

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2008). A Bayesian approach to diffusion process models of decision-making. In V. M. Sloutsky, B. C. Love, & K. McRae (Eds.), *Proceedings of the 30rd annual conference of the cognitive science society* (pp. 1429–1434). Austin, TX: Cognitive Science Society.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response time. *Psychological Methods, 16,* 44–62.

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 Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience, 6,* 1–13.

Vickers, D., & Lee, M. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences, 2,* 169–194.

Vickers, D., & Lee, M. (2000). Dynamic models of simple judgements: II. Properties of a self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multi-choice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences, 4,* 1–31.

Waldorp, L., Christoffels, I., & van de Ven, V. (2011). Effective connectivity of fMRI data using ancestral graph theory: Dealing with missing regions. *NeuroImage, 54,* 2695–2705.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9,* 971–978.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics, 7,* 14.

Winkel, J., van Maanen, L., Ratcliff, R., van der Schaaf, M., Van Schouwenburg, M., Cools, R., & Forstmann, B. U. (2012). Bromocriptine does not alter speed-accuracy tradeoff. *Frontiers in Decision Neuroscience, 6,* 126.

Zandbelt, B. B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences, 111,* 2848–2853.

Received December 9, 2013
Revision received October 9, 2014
Accepted December 9, 2014 ■