On The Efficiency Of Neurally-Informed Cognitive Models To Identify Latent Cognitive States

Guy E. Hawkins[a], Matthias Mittner[b], Birte U. Forstmann[a], and Andrew Heathcote[c]

[a] *Amsterdam Brain and Cognition Center, University of Amsterdam, Amsterdam, The Netherlands*

[b] *Department of Psychology, University of Tromsø, Tromsø, Norway*

[c] *School of Medicine – Division of Psychology, University of Tasmania, Hobart, Tasmania, Australia*

Abstract

Psychological theory is advanced through empirical tests of predictions derived from quantitative cognitive models. As cognitive models are developed and extended they tend to increase in complexity – leading to more precise predictions – which places concomitant demands on the behavioral data used to discriminate between candidate theories. To aid discrimination between cognitive models and, more recently, to constrain parameter estimation, neural data have been used as an adjunct to behavioral data, or as a central stream of information, in the evaluation of cognitive models. Such a model-based neuroscience approach entails many advantages, including precise tests of hypotheses about brain-behavior relationships. There have, however, been few systematic investigations of the capacity for neural data to constrain the recovery of cognitive models. Through the lens of cognitive models of speeded decision-making, we investigated the efficiency of neural data to aid identification of latent cognitive states in models fit to behavioral data. We studied two theoretical frameworks that differed in their assumptions about the composition of the latent generating state. The first assumed that observed performance was generated from a mixture of discrete latent states. The second conceived of the latent state as dynamically varying along a continuous dimension. We used a simulation-based approach to compare recovery of latent data-generating states in neurally-informed versus neurally-uninformed cognitive models. We found that neurally-informed cognitive models were more reliably recovered under a discrete state representation than a continuous dimension representation for medium effect sizes, although recovery was difficult for small sample sizes and moderate noise in neural data. Recovery improved for both representations when a larger effect size differentiated the latent states. We conclude that neural data aids the identification of latent states in cognitive models, but different frameworks for quantitatively informing cognitive models with neural information have different model recovery efficiencies. We provide full worked examples and freely-available code to implement the two theoretical frameworks.

Keywords: Cognitive model, Behavioral data, Neural data, Model recovery, Simulation.

☆This research was supported by a Netherlands Organisation for Scientific Research (NWO) Vidi grant to Birte Forstmann and an Australian Research Council (ARC) Professorial Fellowship to Andrew Heathcote. The authors declare no competing financial interests.

*Correspondence and requests for materials should be addressed to Guy Hawkins, Amsterdam Brain and Cognition Center, University of Amsterdam, Nieuwe Achtergracht 129, Amsterdam 1018 WS, The Netherlands; Email: guy.e.hawkins@gmail.com.

1. Introduction

Quantitative models that explicate the cognitive processes driving observed behavior are becoming increasingly complex, leading to finer-grained predictions for data. Although increasingly precise model predictions are undoubtedly a benefit for the field, they also increase the demands placed on data to discriminate between competing models. The predictions of cognitive models have traditionally been tested against behavioral data, which is typically limited to choices and/or response times. Such behavioral data have been extremely useful in discriminating between model architectures (e.g., Anderson et al., 2004; Brown and Heathcote, 2008; Forstmann et al., 2016; Nosofsky and Palmeri, 1997; Ratcliff and Smith, 2004; Shiffrin and Steyvers, 1997; Tversky and Kahneman, 1992). As model predictions increase in precision, however, we approach a point where behavioral data have limited resolution to further constrain and discriminate between the processes assumed by the models of interest.

The problem of behavioral data providing limited constraint is compounded when one aims to study non-stationarity. Cognitive models typically assume a stationary generative process whereby trials within an experimental condition are treated as independent and identically distributed random samples from a probabilistic model with a specified set of parameters. This assumption has proven extremely useful, both practically and theoretically, but is not supported by fine-grained empirical analysis (e.g., Craigmile et al., 2010; Wagenmakers et al., 2004). Recent work in the study of stimulus-independent thought, or mind wandering, provides a psychological mechanism that can explain these findings, at least in part, in terms of observed performance arising from two or more latent data-generating states. One prominent theory proposes that ongoing performance is driven by two distinct phases: perceptual coupling – where attentional processes are directed to incoming sensory input and completing the ongoing task – and perceptual decoupling – where attention is diverted from sensory information toward inner thoughts (for detailed review, see Smallwood and Schooler, 2015). The perceptual decoupling hypothesis of mind wandering proposes, therefore, that observed behavior is the end result of a mixture of discrete latent data-generating states. To gain insight into the processes underlying the phases of perceptual coupling and decoupling, the goal of the cognitive modeler is to use the available data to determine the optimal partition of trials into latent states.

On the basis of behavioral data alone, such as choices and response times, reliably identifying discrete latent states can be difficult or near impossible. In an example of this approach, Vandekerckhove et al. (2008) aimed to identify *contaminant* trials – data points not generated by the process of interest – in a perceptual decision-making experiment. They defined a latent mixture model in a Bayesian framework that attempted to partition trials that were sampled from the (diffusion model) process of interest from contaminant trials distributed according to some other process. In attempting to segment trials to latent classes, the diffusion model was only informed by the same choice and response time data it was designed to fit. For a representative participant, only 0.6% of their 8000 trials were classified as contaminants, indicating either a remarkable ability of the participant to remain on task (which is unlikely; see, e.g., Killingsworth and Gilbert, 2010), or, more likely, to the limited ability of behavioral data alone to segment trials into latent states.

Rather than relying solely on behavioral data, here we examine whether augmenting cognitive models with an additional stream of information – such as neural data, whether that involves single cell recordings, EEG, MEG, or fMRI – aids identification of latent data-generating states underlying observed behavior. Our aim is to investigate whether the addition of neural data can improve our account of the behavioral data, and in particular the identification of latent states, rather than accounting for the joint distribution of behavioral and neural data (for joint modeling approaches, see Turner et al., 2013a). To this end, we condition on neural data; that is, we do not consider generative models of neural data. Rather, we explore tractable and simple methods that augment cognitive models using neural data as covariates in order to gain greater insight into cognition than is possible through consideration of behavioral data in isolation.

Throughout the manuscript we position our work within the theoretical context of mind wandering. Over the past decade, the scientific study of mind wandering has received great interest from behavioral (e.g., Bastian and Sackur, 2013; Cheyne et al., 2009) and neural (e.g., Andrews-Hanna et al., 2010; Christoff et al., 2009; Weissman et al., 2006) perspectives, though there have been few attempts to integrate the two streams of information in a model-based cognitive neuroscience framework (for an exception, see Mittner et al., 2014). The study of mind wandering is particularly relevant to our aim of identifying latent cognitive states as it is a phenomenon that has been studied under various, qualitatively distinct, hypotheses about how latent states give rise to observed performance (Smallwood and Schooler, 2006, 2015), which we expand upon below. Mind wandering, therefore, serves as an excellent vehicle through which to demonstrate our methodological approach. Our working hypothesis is that mind wandering is a neural state or process that affects the parameters of cognitive models, which in turn affect observed behavioral performance (Hawkins et al., 2015). Our approach inverts this chain of causation: we fit behavioral data with cognitive models that are informed with neural data, and compare their fit to cognitive models that are not informed with neural data. This allows us to assess what can be learnt about mind wandering in a way that is not feasible without the discriminative power of the neural data.

Through the lens of cognitive models of speeded decision-making, we consider two approaches that use neural data to constrain cognitive models, which in turn helps to identify both when people mind wander and the effect it has on task performance. We note, however, that our methods generalize to any domain of study that utilizes neural data – or any additional stream of data, for that matter – to aid identification of latent data-generating states and fit the behavioral data arising from those states with cognitive models.

We consider two general approaches to incorporating mind wandering within a modeling framework. The first approach assumes that observed behavior arises from a mixture of discrete latent states, which may have partially overlapping or unique sets of data-generating parameters. We refer to this as the *Discrete State Representation*. One might think of the latent states as reflecting an *on-task* state, where attention is directed to external stimuli, or task-related thoughts, and an *off-task* state, where attention is directed to internal stimuli, or task-unrelated thoughts, similar to the perceptual decoupling hypothesis (Smallwood and Schooler, 2015). Alternatively, the latent states might reflect *executive control*, where an executive system oversees maintenance of goal-directed behavior, and *executive failure*, which occurs when the executive control system fails to inhibit automatically cued internal thoughts that derail goal-directed behavior (McVay and Kane, 2010). Regardless of the labels assigned to the latent states, models assuming a discrete state representation aim to first identify the mutually exclusive latent states and then estimate partially overlapping or distinct sets of model parameters for the discrete states (for a similar approach, see Mittner et al., 2014). We note that a discrete state representation is also considered outside the context of mind wandering. For example, Borst and Anderson (2015) developed a hidden semi-Markov model approach that used a continuous stream of EEG data to identify discrete stages of processing in associative retrieval.

The second approach generalizes the discrete state representation, relaxing the assumption that latent states are mutually exclusive. This approach assumes a dynamically varying latent state where, for example, at all times a participant will fall at some point along a continuum that spans from a completely on-task focus through to a completely off-task focus. We refer to this second approach as the *Continuous Dimension Representation*, and it approximates 'executive resource' theories of mind wandering (e.g., Smallwood and Schooler, 2006; Teasdale et al., 1995). This class of theories states that executive resources are required to perform goal-directed tasks. The pool of resources is finite, and competing demands, such as mind wandering from the task at hand, reduce the resources available to complete the primary task, leading to suboptimal task performance. The resources available to complete a task can effectively be considered a continuous variable: at times there are more resources available to complete the task than others, and this can vary in potentially complex ways from one trial to the next. Models assuming a continuous dimension representation aim to regress single-trial measures of neural activity onto structured trial-by-trial variation in model parameters (for similar approaches, see Cavanagh et al., 2011; Frank et al., 2015; Nunez et al., 2015, in press). To the extent that the single-trial regressors index the latent construct of interest, this approach dynamically tracks the effect of neural fluctuations on changes in model parameters.

We use a simulation-based approach to explore how well neural data constrains the identification of data-generating states when fitting cognitive models to behavioral data. We first simulate data from models that assume a non-stationary data-generating process (i.e., a latent cognitive state that changes throughout the course of an experiment). We then fit models to the synthetic data that vary in their knowledge of the latent data-generating states: some models completely ignore the presence of a latent mixture in data (i.e., they are misspecified), and others assume partial through to perfect knowledge of the latent data-generating states. The degree of partial knowledge about latent states is assumed to reflect the precision of neural data that informs the analysis. When a neural measure or measures are perfectly predictive of the latent generating states, the partition of behavioral data to one latent state or another mirrors the data-generating process, and the model that assumes a mixture of latent generating states will be preferred over the (misspecified) model that marginalizes over latent states. As the strength of the relationship between the neural measure and the partition in behavioral data weakens, we ought to obtain less evidence for the model that assumes a mixture of latent states in data. Our primary aim is to determine the amount of noise that can be tolerated in the relationship between neural and behavioral data before the misspecified model that collapses across the (true) latent states is preferred. Our outcome measure of interest is, therefore, the probability with which we select the model that assumes more than one latent generating state in data, which was the true data-generating model in all cases.

1.1. Diffusion Model of Speeded Decision-Making

In all simulations we studied sequential sampling models of decision-making, and the diffusion model of speeded decision-making in particular (Forstmann et al., 2016; Ratcliff and McKoon, 2008; Smith and Ratcliff, 2004). The diffusion model, as with most sequential sampling models, assumes that simple decisions are made through a gradual process of accumulating sensory information from the environment. The sensory information influences an evidence counter that tracks support for one response alternative over another; for example, whether a motion stimulus moves to the left or right of a display, or whether a string of letters represents a word or not. The evidence counter continues to track evidence for the two response alternatives until it crosses an absorbing boundary – a pre-determined threshold amount of evidence – which triggers a response. The predicted choice is determined by the boundary that was crossed, and the predicted response time is the time taken for the process to reach the boundary plus a fixed offset time to account for processes such as encoding the stimulus and producing a motor response (e.g., a button press).

Figure 1 provides a schematic overview of a choice between leftward and rightward motion in the diffusion decision model. The model has four core processing parameters: the starting point of evidence accumulation, which can implement biases toward one response or another (z); the average rate at which information is extracted from the stimulus, known as the drift rate (v), the amount of evidence required for a response, which represents cautiousness in responding, known as boundary separation (a); and the time required for elements outside the decision process, known as non-decision time (T_{er}). Modern implementations of the diffusion model assume trial-to-trial variability in some model parameters to reflect the assumption that performance has systematic and nonsystematic components over the course of an experiment (Ratcliff and Tuerlinckx, 2002). These parameters include the drift rate, starting point, and non-decision time. Specifically, on trial i the drift rate is sampled from a Gaussian distribution with mean v and standard deviation η, $v_i \sim N(v, \eta)$; the start point is sampled from a uniform distribution with range sz, $z_i \sim U(z - \frac{sz}{2},\ z + \frac{sz}{2})$; and the non-decision time is sampled from a uniform distribution with range st, $T_{er,i} \sim U(T_{er} - \frac{st}{2},\ T_{er} + \frac{st}{2})$.

In all cases we simulated data from a hypothetical experiment of a two-alternative forced choice task with a single condition. The use of a single experimental condition mirrors almost all laboratory-based studies of mind wandering, which tend to focus on vigilance tasks such as the sustained attention to respond task (SART; Robertson et al., 1997; Smallwood and Schooler, 2006; Smilek et al., 2010). The SART is typically implemented as a single-condition go/no-go task with infrequent no-go stimuli (i.e., stimuli requiring a response to be withheld), with the aim of inducing boredom and hence mind wandering. The sequential sampling models we study here are easily

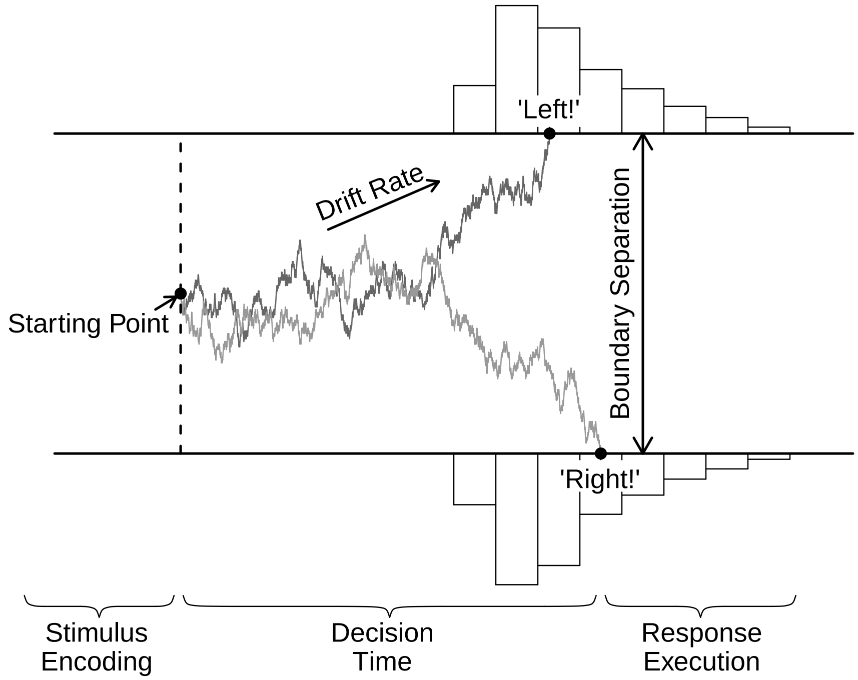

Figure 1: Schematic representation of the diffusion model of speeded-decision making. Reproduced with permission from Hawkins et al. (2015).

generalizable to experimental paradigms with partial response time data – such as go/no-go and stop-signal tasks (Gomez et al., 2007; Logan et al., 2014) – so the results reported here are relevant to the tasks and experimental paradigms typically studied in the mind wandering literature.

Our primary aim was to identify the latent data-generating states in data. This is a question pertinent to the individual-participant level – when was the participant on-task, and when were they off-task – thus we simulate and fit models to data at the individual-participant level.

2. Discrete State Representation

2.1. Generating Synthetic Data

Synthetic data were generated from the discrete state representation by assuming that 80% of trials were from the on-task state and the remaining 20% of trials were from the off-task state. One could manipulate the ratio of on-task to off-task trials as a parameter of the model recovery exercise. We chose instead to select a fixed value that might be considered a conservative estimate of reported rates of mind wandering in experimental tasks that mirror the setup of our simulated experiment, so as to not overstate the estimated power of our results (e.g., some have reported that mind wandering occurs between 30–50% of the time; Killingsworth and Gilbert, 2010).[1]

[1]Nevertheless, to assure ourselves that our results were not dependent on the ratio of on-task to off-task trials and the parameter settings described below, we conducted a parallel analysis where synthetic data were generated from a discrete state representation with

In generating synthetic data we constrained the parameters of the on-task and off-task states to identical values, except for the magnitude of the drift rate. We made the plausible assumption that the drift rate for the on-task state was larger than the drift rate for the off-task state, which implies that mind wandering reduces the efficiency of information processing. This assumption is consistent with empirical results suggesting that mind wandering leads to slower and more variable response times with a greater error rate (e.g., Bastian and Sackur, 2013; Cheyne et al., 2009), which is qualitatively similar to the effect of a reduction in drift rate. Specifically, we set the drift rate for the on-task state to $v_{on} = 2$ and the off-task state to $v_{off} = 1$. All other parameters were set to the following values, for both states: $a = 1$, $z = .5$ (i.e., no response bias), $T_{er} = .15$s, $\eta = 1$, and the trial-to-trial variability parameters for the start point of evidence accumulation and non-decision time were both set to 0. The diffusion coefficient was fixed to $s = 1$ in all synthetic data and model fits were obtained using the 'rtdists' package for the R programming environment (Singmann et al., 2016). An exemplary synthetic data set is shown in Figure 2a and 2b. The synthetic data of the on-task state differed to the off-task state in terms of higher accuracy and faster mean response times that were less variable. These differences indicate that there was a reliable signal in behavioral data that differentiated the latent states.

We generated synthetic data across a wide range of sample sizes (i.e., number of trials completed by a synthetic participant). Our motivation was to determine the efficiency of neural data to identify discrete latent states using sample sizes considered very small for fitting sequential sampling models to data, through to an approximate asymptotic limit with very large sample sizes. Specifically, we simulated 200 synthetic data sets from each of sample sizes 100, 250, 500, 1000, 2000, 5000, and 10000 trials. Therefore, for sample sizes of 100 trials, for example, there were 80 'on-task' and 20 'off-task' trials, and for 10000 trials there were 8000 'on-task' and 2000 'off-task' trials.

2.2. Model Specification

We fit two types of diffusion models to each synthetic data set: a single-state and a dual-state model. In the Appendix we outline the steps involved in performing an analysis assuming a discrete state representation and provide accompanying R code (R Core Team, 2016) that uses the *rtdists* package (Singmann et al., 2016).

2.2.1. Single-State Model

The single-state model is a misspecified model in the sense that it marginalizes (collapses) over trials generated from the on-task and off-task latent states; this approach is equivalent to not using any neural data to inform cognitive modeling. The single-state modeling is representative of the dominant approach in the literature that

an equal ratio of on-task to off-task trials and a lower drift rate for the on-task state ($v_{on} = 1.8$). Following (4) and (5), these settings give an equivalent effect size to that reported in the primary simulation. All results of the parallel analysis mirror those shown in the left panel of Figure 3. Combined with the results shown in Figure 4, this finding suggests that the primary factor influencing recovery of the true latent generating state is the size of the effect that the neural data exert on the latent state, and not particular data-generating parameter settings of the cognitive model.

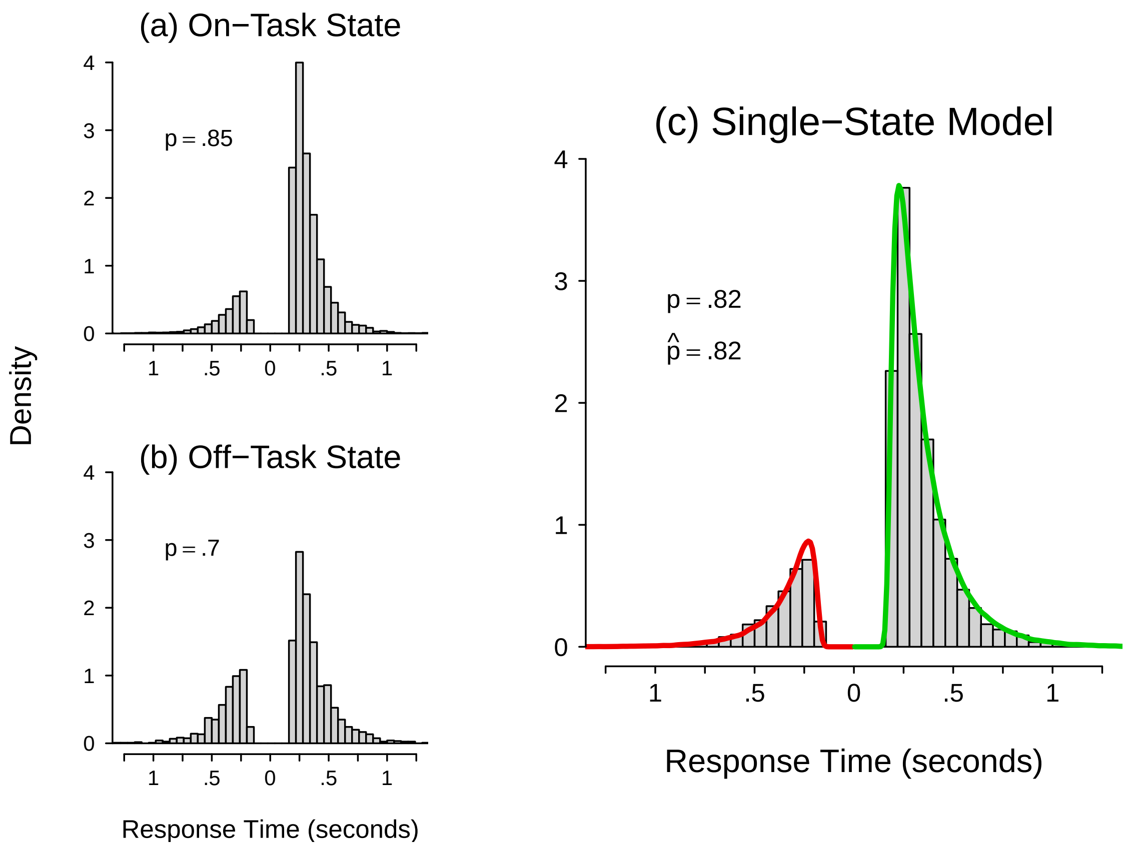

Figure 2: An exemplary synthetic data set generated from the on-task and off-task states of the dual-state model (panels a and b), and the fit of the single-state model to the same data set, collapsed over latent states (panel c). Response time distributions for correct responses are shown to the right of zero and distributions for error responses are shown to the left of zero (i.e., mirrored around the zero-point on the x-axis). Green and red lines show correct and error responses, respectively, from the posterior predictive distribution of the single-state model (panel c). The probability of a correct response in synthetic data is denoted p, and the corresponding predicted probability from the single-state model is denoted $\hat{p}$ (panel c).

generally makes no attempt to account for potential task-unrelated thoughts and their effects on task performance. The single-state model freely estimated the following parameters from data: start point (z), trial-to-trial variability in start point (sz), boundary separation (a), drift rate (v), trial-to-trial variability in drift rate (η), and non-decision time (T_{er}). Trial-to-trial variability in non-decision time was fixed to $st = 0$. We made this decision as we deemed it unlikely that the parameter estimation routine would compensate for the misspecification of the single-state model with a change in the parameter reflecting non-decision time variability, and our Bayesian parameter estimation routines were computationally much more feasible without the numerical integration required for estimation of the st parameter.

2.2.2. Dual-State Model

The dual-state model acknowledged the on-task and off-task generating states in data, by allowing for differences in drift rate between trials allocated to the on-task and off-task states (i.e., freely estimated v_{on} and v_{off}, respectively). All other model parameters were constrained to be equal across the two states (as in the single-state model, $st = 0$ was fixed everywhere). The dual-state model, therefore, assumed some knowledge of the data-generating structure in that there were two states that differed only in drift rate. Our results can thus be interpreted as a 'best case' scenario; additional misspecification in free parameters across the discrete states, or in the number of discrete states, may worsen model recovery relative to the single-state model.

We did, however, introduce misspecification to the dual-state model in terms of the reliability with which trials were allocated to the true generating state. That is, we systematically manipulated the probability that trials generated from the on-task state were in the set of trials allocated to the on-task state in the fitted model, and similarly for the off-task state. In the sense that the set of trials generated from the on-task state was not necessarily the same set of trials fitted as the 'on-task' state, this model is misspecified. We refer to this form of misspecification as *state-level misspecification*, which is distinct from parameter misspecification (i.e., allowing the wrong parameters to vary with state). State-level misspecification mimics the capacity for an external stream of information, such as a neural data, to reliably partition trials into the true (data-generating) latent state. For example, Mittner et al. (2014) trained a support vector machine to use a range of fMRI and pupil measurements to classify trials from a stop-signal paradigm to on-task or off-task states. Their classifier achieved expected accuracy of 79.7% (relative to self-reported mind-wandering), implying that they could expect to correctly classify four out of every five trials to the on-task or off-task states, assuming there was a true distinction in the two latent states in the data-generating process.

Although it is likely that our simulated neural data leads to better-than-chance classification accuracy, no combination of neural measures will achieve 100% accuracy. To explore the effect of classification accuracy on recovery of the (true) dual-state model, we manipulated state-level misspecification in terms of the probability of correctly assigning a trial to its true generating state, which we denote $p_{correct}$. For example, $p_{correct} = .8$ indicates that every trial that was generated from the on-task state had .8 probability of being correctly assigned to the on-task state in the fitted model, and .2 probability of incorrect assignment to the off-task state in the fitted model. The reverse was also assumed: trials generated from the off-task state had .8 probability of assignment to the off-task state in the fitted model, and .2 probability of assignment to the on-task state. This value mimics the classification accuracy achieved in Mittner et al. (2014). We explored a range from $p_{correct} = .5$ (the neural data provide no information about the latent state, so trials are randomly allocated to the on- or off-task state) through to $p_{correct} = 1$ (the neural data provide perfect knowledge of the generating state), in increments of .05. Therefore, for each synthetic data set, we compared the fit of the single-state model to 11 dual-state models corresponding to the range in $p_{correct}$. For each value of $p_{correct}$ we determined which model (single state, dual state) provided the most parsimonious account of the synthetic data set.

2.3. Parameter Estimation

We sampled from the joint posterior distribution of the parameters of each model using differential evolution Markov chain Monte Carlo (Turner et al., 2013b). We assumed prior distributions that had a considerable range around, but conveyed relatively little information about, the true data-generating parameter values:

$$v \text{ [single-state]} \sim N(0,2,-5,5),$$
$$v_{on},\ v_{off} \text{ [dual-state]} \sim N(0,2,-5,5),$$
$$a,\ sv \sim N(1,1,0,2),$$
$$z,\ sz,\ T_{er} \sim Beta(1,1),$$

where $N(\mu,\sigma,a,b)$ denotes a Normal distribution with mean μ, standard deviation σ, truncated to a lower limit of a and upper limit of b, and $Beta(\alpha,\beta)$ denotes the Beta distribution with shape parameters α and β. Parameters z and sz were estimated as a proportion of parameter a, and hence were constrained to the unit interval.

Independently for all models, we initialized 18 chains with random samples from the prior distribution. Chains were first run for 250 iterations with the differential evolution probability of migration set to .05. Once initialization was complete, the migration probability was set to zero and we sampled from the joint posterior distribution of the parameters in phases of 1000 iterations. After each phase we checked chain convergence using the multivariate potential scale reduction factor ($\hat{R}$ statistic; Brooks and Gelman, 1998), using a criterion of $\hat{R} < 1.15$ to indicate convergence (visual inspection of a sample of chains supported this conclusion).[2] After each phase of 1000 iterations we monitored whether the chains had converged. If so, the parameter estimation routine was terminated. If not, another 1000 iterations were started from the end point of the previous 1000 iterations, and the procedure repeated until the chains had converged.

2.4. Model Selection

Model selection was performed with the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002)[3], which is computed using samples from the joint posterior parameter distribution. DIC is defined as DIC $= D(\overline{\theta}) + 2p_D$, where $D(\overline{\theta})$ is the deviance at the mean of the sampled posterior parameter vector θ, and p_D is the effective number of model parameters, where $p_D = \overline{D} - D(\overline{\theta})$, and $\overline{D}$ is the mean of the sampled posterior parameter deviance values. Lower values of DIC indicate the better model for the data (i.e., the most parsimonious tradeoff between goodness of fit and model complexity).

We converted estimated DICs for each comparison of the single- and dual-state models to model weights (for overview, see Wagenmakers and Farrell, 2004). If the set of models under consideration contain the true data-generating model, then these weights provide estimates of the posterior probability of each model (i.e., the

[2]Preliminary simulations indicated lower values of $\hat{R}$ (e.g,. $\hat{R} < 1.1$) were produced by longer series, but without any change in conclusions; we chose a length of 1000 as a compromise that kept computational demands feasible.

[3]DIC has been criticized because it can select models that are too complex. Gelman et al. (2014) favor instead an information criterion that approximates Bayesian leave-one-out cross validation, WAIC (Watanabe, 2013); for a number of checks we performed on our extensive simulation study DIC and WAIC produced almost identical results. The code we provide to apply our analyses allows calculation of both information criteria, so users can use their preferred choice.

probability conditional on the data of each model being the true model relative to the set of candidate models under comparison). Otherwise, model weights provide a graded measure of evidence rather than the all-or-none decision rule that can arise when interpreting 'raw' information criteria. Model weights are also on the same scale for different data-set sizes (i.e., they fall on the unit interval), which allowed for simple comparison of model recovery across the sample sizes that were systematically manipulated in our study.

Model weights are calculated by first considering differences in DIC for each model fit to a given data set: $\Delta_i(\text{DIC}) = \text{DIC}_i - \min \text{DIC}$, where min DIC is the lowest (i.e., best) DIC among the set of K models under consideration. Then, the DIC-based weight for model i, $w_i(\text{DIC})$, from the set of K models is given as

$$w_i(\text{DIC}) = \frac{\exp\Big\{-\frac{1}{2}\Delta_i\big(\text{DIC}\big)\Big\}}{\sum_{k=1}^{K}\exp\Big\{-\frac{1}{2}\Delta_k\big(\text{DIC}\big)\Big\}}. \tag{1}$$

We calculated model weights for pairwise comparisons between the single- and dual-state models. All synthetic data were generated from the dual-state model so our primary outcome measure was the weight in favor of the dual-state model (i.e., successful model recovery), given by a simplified form of Equation 1,

$$w_{dual}(\text{DIC}) = \frac{\exp\Big\{-\frac{1}{2}\Delta_{dual}\big(\text{DIC}\big)\Big\}}{\exp\Big\{-\frac{1}{2}\Delta_{single}\big(\text{DIC}\big)\Big\} + \exp\Big\{-\frac{1}{2}\Delta_{dual}\big(\text{DIC}\big)\Big\}}. \tag{2}$$

We calculated model weights according to (2) for all relevant comparisons, and then averaged over the 200 Monte Carlo replicates within each state-level misspecification (.5, .55, ..., .95, 1) by sample size (100, 250, 500, 1000, 2000, 5000, 10000) cell of the design.

2.5. Results and Discussion

The single- and dual-state models provided an excellent fit to all synthetic data sets. Figure 2c shows the fit of the single-state model to an exemplary synthetic data set. It is perhaps surprising, but also instructive, that the misspecified single-state model provided such a precise account of data generated from two discrete latent states that had different data-generating parameters. It appears that the single-state model is able to mimic the dual-state model, at least for the parameter settings we investigated. Specifically, when the drift rate is the only parameter that varies across discrete states – where v_{on} and v_{off}, respectively, represent drift rates for the on-task and off-task states, and p_{on} represents the proportion of on-task trials – the estimated (single) drift rate of the misspecified single-state model approximates a weighted combination of the two: $v_{on} \times p_{on} + v_{off} \times (1 - p_{on})$. To mimic the variability of the mixture of drift rate distributions – which is increasingly greater than the variability of either of the mixture components as the two means increasingly differ – there is an increase in the standard deviation of the trial-to-trial variability in drift rate (η) estimate for the single-state model. For the difference in drift rates that we investigated this increase was only marginal, and the slightly more variable single drift rate distribution approximated the mixture distribution quite well (see also discussion around formulae (4) and (5) below). This approximation will likely break down as the difference in means becomes extreme, but as the difference we examined was quite substantial it seems unlikely that visual examination of goodness-of-fit alone would be sufficient in practice to detect a misspecified single-state model.

Since both models provided a visually compelling fit to behavioral data, we discriminated between the single- and dual-state models on the basis of model weights, as is standard in most research comparing competing cognitive models. The left panel of Figure 3 summarizes the model recovery simulation. The weight in favor of the dual-state model – the true data-generating model – is shown on the y-axis. Light through to dark lines indicate the amount of state-level misspecification, where classification to the true latent state was manipulated from chance performance ($p_{correct} = .5$, lightest line) through to perfect classification ($p_{correct} = 1$, darkest line). The key comparison is the ability to identify the true latent generating state on the basis of cognitive models fit to behavioral data, across a range of neurally-informed classification accuracies.

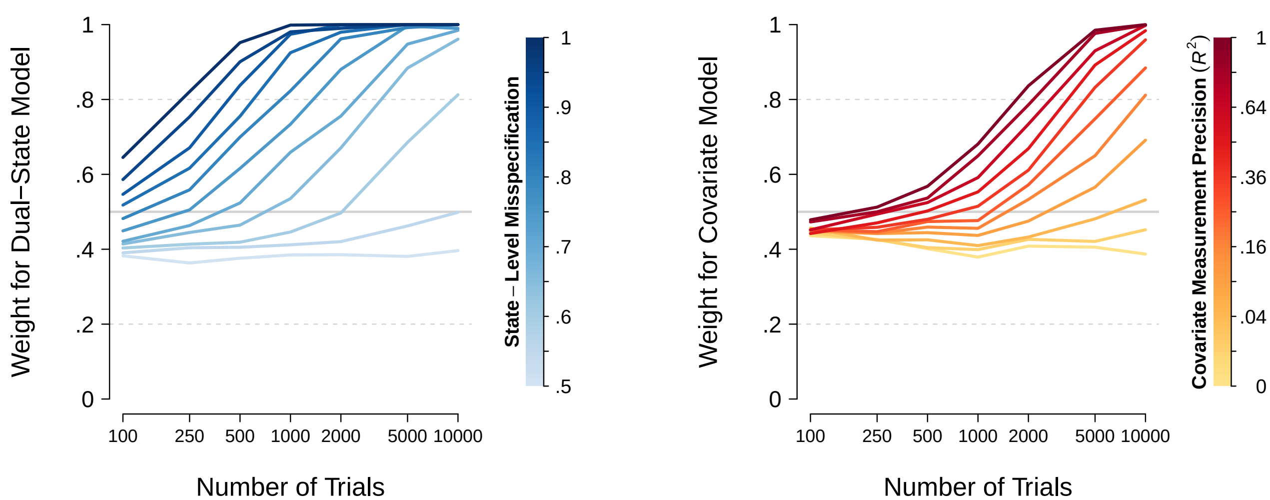

Figure 3: Model recovery for medium effect sizes. The left panel shows the weight in favor of the dual-state model over the single-state model in the model recovery simulations of the discrete state representation. The y-axis represents the DIC-derived posterior model probability of the dual-state model, the x-axis represents the number of trials in the synthetic data set, and color gradations represent the range in $p_{correct}$ of the state-level misspecification of the dual-state model. The right panel shows the weight in favor of the covariate model over the standard model in the model recovery simulations of the continuous dimension representation. The y-axis represents the DIC-derived posterior model probability of the covariate model and color gradations represent the range in R^2 of the covariate measurement precision of the covariate model. Horizontal gray lines indicate the point of equivalent evidence between the two models (solid lines), and a difference of approximately 3 DIC units in favor of the dual-state model (left) and covariate model (right; upper dashed lines) or the single-state model (left) and standard model (right; lower dashed lines).

As expected, evidence in favor of the dual-state model increased as the number of trials in the synthetic data increased (larger values on the x-axis). This was, however, heavily influenced by the amount of state-level misspecification. In our simulations, this represents the capacity of the neural data to reliably classify trials to their true latent (data-generating) state. Whenever state-level misspecification was above chance (i.e., $p_{correct} > .5$), the evidence in favor of the dual-state model increased with increasing sample size. In particular, it reached ceiling by a sample size of 1000 trials when state-level misspecification was completely absent ($p_{correct} = 1$), and by the upper limit of the sample sizes we explored (10000 trials) for moderate classification accuracy ($p_{correct} \geq .7$). For more plausible sample sizes, however, recovery of the true model was more modest. Even with no state-level misspecification, the weight for the dual-state model never exceeded .8 for sample sizes less than 250 trials. We note that a model weight of .8 corresponds to a difference of approximately 3 units on the raw DIC scale. Small differences in information criteria such as this are often considered as providing little more than weak evidence (e.g., Burnham and Anderson, 2004; Kass and Raftery, 1995; Raftery, 1995). Even placing optimistic bounds on the level of classification accuracy that is possible with real neural data (e.g., $p_{correct} = .9$), the weight for the dual-state model only exceeded .8 at a sample size of approximately 400 trials, and did not reach a decisive level of evidence until the sample size exceeded 1000 trials.

On a more technical point, when state-level misspecification was at chance ($p_{correct} = .5$), the single-state model ideally ought to garner increasing evidence with increasing sample size (i.e., a gradual shift toward lower values on the y-axis). This should occur since the classification to discrete states in the fitted model was completely uninformed by the true data-generating values, so the estimated drift rates for trials classified to the on- and off-task states were close to identical. Under these conditions, the dual-state model provides no predictive benefit over the single-state model, so we should favor the simpler single-state model, and increasingly so for larger sample sizes. Examination of Figure 3, however, indicates that this did not occur; model weight was independent of sample size. This result is due to a property of the model selection criteria used here. DIC penalizes model complexity with a fixed offset (the effective number of parameters, p_D), which means that the penalty against the dual-state model over the single-state model when $p_{correct} = .5$ is (almost) a fixed value as a function of the sample size manipulation in our study, hence the approximately flat line at $y = .4$. This problem would be addressed through the use of model selection indices that are consistent in the sense that they converge to the true answer with increasing sample size, such as Bayes factors. At the time of this work, calculation of Bayes factors for complex cognitive models such as the diffusion model is computationally extremely expensive. This is an active field of research and with future developments we hope to incorporate such model selection measures in our work (for a recent example, see Steingroever et al., 2016).

In summary, our simulation study indicates that it can be difficult to identify discrete latent states on the basis of cognitive models fit to behavioral data. Of course, it is possible that changes to the parameters of the simulation may alter these results. For example, we could manipulate the ratio of on-task to off-task trials in synthetic data, the number of model parameters that differed across the latent states and the degree of difference, or the level of parameter misspecification in the models fit to the synthetic data. On the basis of the available evidence, however, we conclude that obtaining compelling evidence for the identification of mutually exclusive latent states – such as phases of on-task and off-task performance – requires very large sample sizes (5000+ trials) with moderate (or better) neural classifiers, or moderate (or better) sample sizes with very good neural classifiers. Our intuition is that neither of these situations arise in the majority of real psychological or neuroscience experiments. Nevertheless, for almost all sample sizes we obtained at least some evidence in favor of the true model for plausible sample sizes (e.g., a few hundred to a few thousand trials) when data were partitioned to discrete states on the basis of neural classifiers that performed within an impressive but plausible range for real data (e.g., $p_{correct} = .7 - .85$).

3. Continuous Dimension Representation

The first model recovery analysis indicated that identifying discrete latent states on the basis of cognitive models fit to behavioral data is difficult but not impractical. We now investigate a generalization of the discrete state representation that considers the latent state as a continuous dimension. In the context of mind wandering, such a continuum could represent a dynamically fluctuating state where people drift into phases of more on-task or more off-task focus, without imposing a rigid boundary between mutually exclusive states. The idea underlying the continuous dimension representation is more general, though, mirroring constructs in many cognitive theories, such as the graded memorability of different items in a recognition memory experiment. Indeed, it was to account for just such graded variability that Ratcliff (1978) introduced trial-to-trial variability in drift rates into the diffusion model, which has since become a standard assumption (i.e., $\eta > 0$).

The continuous dimension representation can be interpreted in two ways. The first assumes that there is an external stream of information, which we assume throughout to be some form of neural data, that reliably indexes a latent state, such as mind wandering. In the mind wandering literature, for example, measures of connectivity and activity of the default mode network are increased during phases of reduced attention toward the primary task (e.g., Andrews-Hanna et al., 2010; Christoff et al., 2009; Mason et al., 2007; Mittner et al., 2014; for meta-analysis, see Fox et al., 2015). In this case, moment-to-moment fluctuations in activity of the default mode network could be considered an online index of mind wandering. This stream of neural data can then be used as a covariate in the cognitive model; specifically, single-trial measures of default mode network activity can be regressed onto structured trial-by-trial variation in the parameters of the model. This allows exploration of the effect of the neural covariate on different model parameters and permits quantitative tests of the covariate-parameter pairings that provide the best fit to behavioral data. This approach has the potential to provide insights regarding how the latent state (e.g., mind wandering as indexed by activity of the default mode network) affects cognition (e.g., processing efficiency; drift rate) and consequent task performance (e.g., more errors, slower response times).

The second way to interpret a continuous dimension is that the neural measure provides a direct 'readout' of a process assumed in the cognitive model. This approach allows for precise tests of 'linking propositions' (Schall, 2004); explicit hypotheses about the nature of the mapping from particular neural states to particular cognitive states. As an example of this approach, Cavanagh et al. (2011) proposed that response caution in conflict tasks is modulated by connectivity between the subthalamic nucleus and medial prefrontal cortex. To test this hypothesis, the authors first estimated single-trial measures of theta band power from neural oscillations in ongoing EEG activity over the medial prefrontal cortex, which was then regressed onto the value of the decision boundary parameter of the diffusion model. This single-trial regressor approach estimates regression coefficients that indicate the valence and magnitude of the relationship between the neural measure and observed performance, via the architecture of the cognitive model. Cavanagh et al. (2011) found that increased theta power led to a subsequent increase in the decision boundary (i.e., a positive value of the regression coefficient) for trials with high but not low conflict. A control analysis indicated that theta power had no trial-level relationship with drift rate (i.e., a regression coefficient centered at zero), indicating a selective effect of the neural measure on a model parameter. This example highlights how single-trial regression permits quantitative tests of hypotheses about brain-behavior relationships.

Regressing neural data onto the parameters of cognitive models at the single-trial level has the desirable property that it provides a tight quantitative link between neural and behavioral data (de Hollander et al., 2016). Furthermore, although we used custom scripts for all analyses reported here – because we needed to automate a large number of replications – there are excellent, freely available programs that implement single-trial regression for hierarchical and non-hierarchical Bayesian parameter estimation for the diffusion model (HDDM toolbox for Python; Wiecki et al., 2013), which removes barriers to implementation of these methods. In the Appendix we outline the steps involved in performing single-trial regression and provide accompanying R code to implement these steps.

In this section we assessed whether the trial-by-trial influence of an external stream of information, such as a neural measure, is identifiable in models fit to behavioral data. In previous simulation studies, Wiecki et al. (2013) found that single-trial covariates are well recovered in a hierarchical estimation setting for moderate effects sizes and moderate number of trials in the experiment. We build on Wiecki et al.'s findings to explore how often a model that incorporates a single-trial neural covariate – which was the true model in all cases – was preferred over the 'standard' diffusion model that uses no trial-level covariates.

3.1. Generating Synthetic Data

Synthetic data were generated from a diffusion model where a neural signal modulated individual-trial drift rates: trials with larger-than-average neural signals had larger-than-average drift rates and trials with smaller-than-average neural signals had smaller-than-average drift rates. We assumed that the neural covariate would be pre-processed and normalized prior to modeling. To this end, we simulated a single value of the neural covariate for every synthetic trial via random draws from the standard normal distribution and explored the effect of the neural covariate on recovery of the data-generating model.

3.1.1. Covariate Model

Synthetic data were generated data from a model that assumed trial-to-trial variability in drift rate had systematic fluctuations, via the neural covariate, and unsystematic (random) fluctuations, via parameter η, which we refer to as the *Covariate* model. We assumed that the trial-level neural covariate was mapped via simple linear regression to structured trial-by-trial variation in drift rate. Specifically, drift rates were distributed according to the value of the normalized covariate (d) and a regression coefficient (β), such that the drift rate (v) on trial i is:

$$v_i \sim v + \beta \cdot d_i + N(0,\ \eta). \tag{3}$$

The covariate model thus assumed that the drift rate on trial i, v_i, had a mean component defined as a linear function of an intercept, v, representing average performance in the experiment, and the magnitude and valence of the neural measure on trial i, d_i, scaled by a regression coefficient, β, which is an index of effect size, and a random component involving samples from a Gaussian distribution with mean 0 and standard deviation η. This model reflects the plausible assumption that our measured neural covariate has a generative influence on drift rate (through parameter β), but there are also unmeasured, randomly distributed influences on drift rate (through parameter η).

3.1.2. Effect Size of the Neural Covariate

We matched the effect size (β) studied in the continuous dimension representation to the effect size studied in the discrete state simulations in terms of the proportion of variance accounted for by the neural information. Specifically, if p_{on} represents the proportion of on-task trials in the discrete state representation, and x_1 and x_2 respectively represent sampled drift rates of the on-task and off-task states, where $x_1 \sim N(v_{on}, \eta_{on})$ and $x_2 \sim N(v_{off}, \eta_{off})$, then the weighted mean drift rate of the mixture is

$$M_{discrete} = p_{on} \cdot v_{on} + (1 - p_{on}) \cdot v_{off}, \tag{4}$$

with variance

$$V_{discrete} = p_{on} \cdot \eta_{on}^2 + (1 - p_{on}) \cdot \eta_{off}^2 + p_{on} \cdot (v_{on} - M_{discrete})^2 + (1 - p_{on}) \cdot (v_{off} - M_{discrete})^2. \tag{5}$$

Substituting the values used in the discrete state simulations ($p_{on} = .8$, $v_{on} = 2$, $v_{off} = 1$, and $\eta_{on} = \eta_{off} = 1$) into (4) and (5) we get $M_{discrete} = 1.8$ and $V_{discrete} = 1.16$. The proportion of variance accounted for by the neural data in the discrete state simulations was therefore

$$R^2_{discrete} = 1 - \frac{1}{V_{discrete}} = 1 - \frac{1}{1.16} = .138,$$

which gives the medium effect size of $r_{discrete} = \sqrt{R^2_{discrete}} = .371$.

We used a comparable definition of effect size for the continuous dimension representation. If the neural data is distributed as $d \sim N(0, V_{neural})$ with regression coefficient β and base drift rate variability $x \sim N(0, \eta)$,[4]

[4] Here we set $V_{neural} = 1$ without loss of generality and similarly both means at zero as we are only concerned with proportions of variance.

then it follows that the covariate model in (3) has variance

$$V_{continuous} = \eta + \beta \cdot V_{neural},$$

with proportion variance

$$R^2_{continuous} = \frac{\beta \cdot V_{neural}}{\eta + \beta \cdot V_{neural}}. \quad (6)$$

Rearranging (6) and setting $R^2_{continuous} = R^2_{discrete} = .138$, we get

$$\beta = \frac{\eta \cdot R^2_{continuous}}{V_{neural}(1 - R^2_{continuous})} = .16,$$

which is the value of the regression coefficient we used to generate synthetic data. This value is broadly representative of the few previous studies that have reported single-trial regression coefficients in empirical studies using a model-based neuroscience framework; $\beta \approx .20$ for drift rate effects in Nunez et al. (in press), and $\beta \approx .09$ and .04 for response threshold effects in Cavanagh et al. (2011) and Frank et al. (2015), respectively. All other parameters of the covariate model were set to the same values as in the simulation of the on-task state of the discrete representation.

We again generated synthetic data sets from the same range of sample sizes as in the previous analysis; 200 synthetic data sets from the covariate model for each of sample sizes 100, 250, 500, 1000, 2000, 5000, and 10000 trials.

3.2. Model Specification

We fit two types of diffusion models to each synthetic data set: the covariate model and a 'standard' model. The covariate model was fit to all synthetic data sets with the drift rate assumptions specified in (3). The second model neglected the information contained in the neural covariate altogether, instead attributing trial-to-trial variability in drift rate to unsystematic sources via the η parameter; that is,

$$v_i \sim N(v,\ \eta).$$

We refer to this second model as the *Standard* model, reflecting its dominant status in the literature (Ratcliff, 1978; Ratcliff and McKoon, 2008).

When the neural signal is measured with perfect precision, the true latent data-generating model – the covariate model – should be favored over the standard model. Such high measurement precision, however, is not possible in real neural data. To examine the effect of noisy neural data on the identification of a model incorporating a neural covariate, we manipulated the level of noise in the covariate that was fit to the synthetic data. That is, we systematically diminished the correlation between the data-generating value of the covariate and the fitted value of the covariate, which we refer to as *covariate measurement precision*. This manipulation mimics the setup of real experiments where we (aim to) obtain neural measures that are noise-perturbed proxies to the true neural state.

To systematically manipulate covariate measurement precision, for each synthetic data set we generated a new set of random variables that served as the neural covariate in the models that were fit to the synthetic data. The set of random variables, which we refer to as 'fitted covariates', had correlations with the data-generating value of the covariate ranging from $r = 0 - 1$ in increments of .1. The mean (zero), variance (one) and shape (normal) of the fitted covariates were the same as that of the covariate distribution.[5]

We report covariate measurement precision below as the coefficient of determination (R^2) rather than Pearson correlation coefficient (r). This allows for direct interpretation as the proportion of variance that the noise-perturbed, fitted value of the covariate accounts for in the true data-generating value of the neural covariate. These results provide a benchmark for the minimum level of measurement precision required for identifiability of cognitive models that incorporate single-trial covariates.

3.3. Parameter Estimation and Model Selection

We estimated model parameters using identical methods to those described in the analysis of the discrete state representation, with the only addition that we specified a prior distribution for the covariate parameter of the covariate model: $N(0, 1, -3, 3)$.

Model selection was also conducted in a parallel manner to the first analysis. Our primary aim was to determine the covariate measurement precision required to obtain evidence in favor of the covariate model over the standard model. Therefore, we report the model weight in favor of data generated from the covariate (i.e., true) model over the standard model, following (2).

3.4. Results and Discussion

All models provided an excellent fit to synthetic data so we again adjudicated between them using model weights. The right panel of Figure 3 summarizes model recovery in a similar format to the left panel. Larger values on the y-axis indicate more evidence for the true (covariate) model over the standard model. Line darkness indicates the level of covariate measurement precision, where measurement precision was manipulated from complete noise ($R^2 = 0$, lightest line) through to perfect measurement ($R^2 = 1$, darkest line). As before, the key comparison was

[5]Under this model of measurement noise, the relationship to the proportion of variance in drift rates explained by mind wandering is more transparent than in the discrete case where measurement noise is in terms of the proportion of correct classifications. To see this, denote the proportion of variance in the measured covariate (MC) by w, and the random variables representing the systematic effect of the covariate and measurement noise by $D \sim N(0, 1)$ and $M \sim N(0, 1)$, respectively. Hence, $MC = w \cdot D + (1 - w) \cdot M$, and so $MC \sim N(0, 1)$ as required. Consequently, the overall drift rate random variable with the measured covariate is $V \sim v + \beta \cdot MC = v + \beta \cdot D + N(0, \sqrt{1 + \beta \cdot (1 - w)})$. These results show the additive Gaussian assumption causes the difference between measurement noise and the random effects on the drift rate unrelated to the covariate not to be identifiable, with the combination constituting what might be called the "effective" level of noise. Given, $r = \sqrt{w}$, our manipulation of r is a manipulation of the effective noise level, corresponding either to a change in the level of measurement noise, the level of unrelated effects on drift rates, or some combination. We maintain the distinction between the two constituents of effective noise in our description of results given it makes clear the link to the discrete case, where in both cases the range of the measurement noise manipulation is between no effect and the maximal effect size (i.e., $.138 = \beta/(1 + \beta)$, where $\beta = .16$).

the capacity to identify the true generating model in neurally-informed versus neurally-uninformed cognitive models fit to behavioral data.

Evidence in favor of the true model generally increased with the number of trials in synthetic data. As expected, however, this was influenced by the level of covariate measurement precision. When the covariate was measured with very low precision – where the fitted value of the covariate explained less than 5% of the variation in the data-generating covariate – sample size had almost no influence on recovery of the true model. This implies that when neural data are poorly measured, or when the neural measure is only a very weak proxy to the true latent process, then a binary decision would select the standard model over a neurally-informed model. That is, assuming unsystematic across-trial variation in drift rate would be more parsimonious than regressing an overly noisy neural measure onto drift rate.

Perhaps surprisingly, evidence for the true model converged very slowly as a function of sample size. Even when the neural covariate was perfectly measured ($R^2 = 1$), the weight for the true model did not exceed .8 until almost 2000 trials were observed; the comparable sample size for the discrete state simulation was 250 trials. For more plausible measurement precision – say, approximately 33% – the weight for the true model exceeded .8 only when sample size exceeded approximately 4000 trials. This result, and similar comparisons across the panels of Figure 3, suggests that the discrete state approach is a more powerful use of neural data than the single-trial covariate approach, at least for the parameter settings and effect size explored here. That is, neural data more heavily constrain model recovery when used as a binary indicator of the latent state than when regressed onto trial-by-trial variation in model parameters.

4. Recovering Neurally-Informed Cognitive Models When Neural Data Have A Large Effect Size

The foregoing analyses indicate that when equated on a medium effect size, neurally-informed discrete state models are more reliably recovered than neurally-informed continuous dimension models. In this section we confirm that when endowed with a sufficiently large effect size the true model is well recovered in both the discrete state and continuous dimension representations. This result implies that both discrete and continuous representations can indeed be identified in behavioral data when the information contained in neural data relates to a sufficiently strong effect.

We generated synthetic data sets where the neural data strongly identified the latent state. Specifically, for the continuous dimension representation we set the value of the neural covariate to $\beta = .5$, with all other parameter settings as described in the previous section. Following (6) this gives an effect size of $r = .577$. An equivalent effect size can be obtained in the discrete state representation in multiple ways. We chose to enhance the difference between the on- and off-task states in terms of a larger drift rate for the on-task state ($v_{on} = 2.414$) and assuming an equal ratio of on-task to off-task trials ($p_{on} = .5$), with no changes in other data-generating parameters. All other details were identical to those used in the previous simulations, including the data generation, sample size, introduction of noise to the (synthetic) neural data, model specification, parameter estimation, and model selection methods.

Figure 4 shows recovery of the true model with a large effect size in the discrete state and continuous dimension representations. A striking finding was how quickly the evidence for the true models converged as a function of the noise in neural data (state-level misspecification and covariate measurement precision in the left and right panels, respectively), even at relatively small samples sizes (i.e., 250–500 trials) and moderate levels of noise. Although recovery of the continuous dimension representation was much improved for large versus medium effect sizes, the true model in the discrete state representation was still recovered more reliably when equating sample size and noise in neural data.

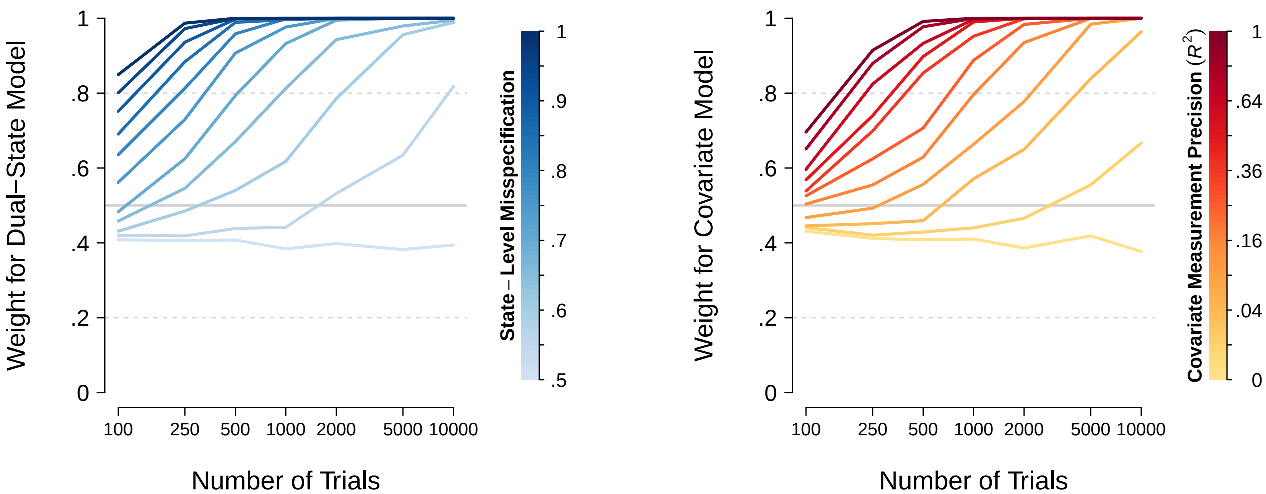

Figure 4: Model recovery for large effect sizes. The left panel shows the weight in favor of the dual-state model over the single-state model for the discrete state representation. The right panel shows the weight in favor of the covariate model over the standard model for the continuous dimension representation. All other details are as described in Figure 3.

5. Conclusions

We investigated whether informing cognitive models with neural data improves the ability to identify latent cognitive states. This approach is increasingly common in the psychology and neuroscience literatures (e.g., Borst and Anderson, 2015; Mittner et al., 2014; Turner et al., 2013a, 2015). However, there have been few systematic studies of the benefits to model recovery that such an approach may bear. We found that, when the neural data can discriminate a moderate effect on performance, it can be difficult to reliably identify mutually exclusive latent states when neurally-informed cognitive models are applied to behavioral data. As expected, model recovery was very good when the synthetic experimental design was far removed from typical experiments (i.e., large sample size, good neural classification accuracy). Model recovery, however, was still within acceptable bounds even with more feasible experimental designs (i.e., between 500-1000 trials) with moderate classification accuracy.

In contrast, when we relaxed the assumption that latent states are discrete, we found that a latent state that can dynamically move along a continuous dimension substantially worsened model recovery even though mind wandering accounted for the same proportion of variance (i.e., had the same effect size) in the continuous and discrete versions. Model recovery was relatively poor for the sample sizes typically observed in psychological experiments (i.e., up to 1000 trials per participant), and convincing evidence for the true data-generating model was only obtained with sample sizes of approximately 5000 trials or more, even when neural covariates were (hypothetically) measured with perfect precision. This result implies that when the neural covariate is only a distant proxy to the true data-generating process, a standard model that is ignorant with respect to neural data will often be preferred over a neurally-informed cognitive model, and, within reason, this is not dependent on sample size. We believe this highlights two important issues in the use of neurally-informed cognitive models. The first, more obvious issue is that we must maximize the precision in our measurement of neural data. **Birte: Following Andrew's comment, is there anything you can add to accompany the previous sentence about precision of neural recording?** The second, more subtle issue is that we must use theory-based, hypothesis-driven tests of neural covariates on model parameters; that is, we must aim to maximize the possible relationship between the fitted value of the covariate and the true data-generating process.

Our conservative conclusion is, therefore, that neural data aids model identification under some circumstances. In particular, model recovery improved when the latent state was assumed to consist of discrete stages (vs. continuous dimension). The discrete approach had greater power in the sense that a given effect could be identified with smaller sample sizes, reflecting more efficient use of neural data. This finding may be due to the parameter governing trial-to-trial variability in drift rate (η) having a better capacity to compensate for variance arising from the neural covariate under the assumption of a continuous dimension than a discrete state representation. Nevertheless, in practice this finding is particularly important since experiments that record neural measures such as fMRI or EEG activity during task completion are often limited in the number of trials that can be collected. Reassuringly, when the neural data exerted a large effect on behavior (although not such a large effect as to be implausible at least in some circumstances) both the discrete state and continuous dimension representations had good model recovery. Even under this condition, however, relative to the standard model, the assumption of mutually exclusive latent data-generating states was more efficiently recovered than a latent continuous dimension (cf. Figure 4). Finally, we note that efficiency of model recovery appears to more heavily influenced by the effect size rather than particular hyperparameter settings (cf. footnote 1).

It is also important to note that even in the large effect size case simple visual inspection of model fits was not sufficient to reject the standard model; we required model selection methods. Fortunately, methods that are easily implemented based on standard Bayesian posterior sampling (e.g., DIC, WAIC) sufficed for detecting the presence of an effect of mind wandering in our simulations. However, more sophisticated model selection methods (e.g., Bayes factors) appear to be required to provide consistent evidence (i.e., evidence that becomes stronger as sample size increases) against the presence of mind wandering. That is, when mind wandering is not present, at best the model selection methods explored here will be equivocal even with large samples.

Regardless of whether one expects neural data to exert a small or large effect on performance, the assumption of a discrete or continuous representation will likely better serve different research goals at different times. Both approaches allow estimation of effect size (cf. formulae 4–6). The continuous approach also has the attractive property that a measure of effect size is directly estimated. That is, the output of the neural covariate-model parameter relationship – a regression coefficient – has a simple interpretation (assuming that the neural covariate is normalized): the extent to which the estimated regression coefficient differs to zero provides a standardized measure of effect size. Both approaches are also relatively easy to implement. The discrete state representation can be implemented by splitting an experimental condition into discrete sets of trials on the basis of a neural variable (e.g., the output of a classifier). Single-trial covariates are already incorporated as a standard feature of some estimation programs (e.g., HDDM, Wiecki et al., 2013), removing a potential barrier to implementation. In the Appendix we also provide custom R code to implement both analysis approaches discussed in this paper.

Our analyses examined recovery of latent cognitive states in individual (simulated) participants, though one could also consider recovery of latent states across groups of participants. This could be investigated with hierarchical Bayesian models that, among other benefits, allow for simultaneous analysis at the level of individuals and groups (for an overview, see Lee, 2011). Such an approach allows information to be pooled across participants in a theoretically sensible manner, which can confer benefits to parameter estimation, in particular parameter stability. Furthermore, hierarchical Bayesian modelling can be applied to large samples of participants, where each participant may only complete a moderate number of trials. However, it is important to note that if there are too few data for each participant then individual differences cannot be estimated, with hierarchical models often displaying "over-shrinkage" (i.e., estimating the same parameter value for all participants). For simplicity, we restricted our analyses to the simpler case of recovering latent cognitive states in individual participants, which removes at least some sources of variability that are present in the hierarchical case (e.g., across-participant variability in the proportion of trials from each of two discrete latent states). We leave these interesting questions about model recovery in hierarchical settings to future research.

Finally, we note that the discrete and continuous approaches need not involve neural data, although we considered such hypothetical scenarios here. A variable derived at the level of single trials – which can be incorporated within a discrete or continuous approach – can be extracted from any property of the task environment that is relevant to performance. For example, Hawkins et al. (2016) studied the similarity between study and test items in an inductive reasoning task. The similarity relations are specified at the level of individual items, and thus can be regressed against parameters of the cognitive model in the same manner as neural data. In Hawkins et al.'s model, regressing single-trial item similarity onto the drift rate parameter led to a positive regression coefficient, indicating that as item similarity increased so too did the probability of generalizing a target property to novel items according to a particular functional form. This example illustrates the general point that wider incorporation of single-trial properties of the experiment – neural or otherwise – in cognitive models has the potential to provide deeper insight to a broad range of psychological phenomena.

Appendix A. Implementing Neurally-Informed Cognitive Models

In this Appendix we outline the steps involved in implementing the discrete state and continuous dimension representations discussed in the main text. To accompany the examples, we provide code in the R programming language (R Core Team, 2016) that is freely available on the Open Science Framework (`osf.io/yt8q4`).

This outline provides guidance on the cognitive modeling component of a model-based neuroscience analysis in real data. It assumes that the neural data – whether it is fMRI, EEG, MEG, pupil measurements, or others – have been analyzed in an appropriate manner. It further assumes that it is possible to extract at least one value of the analyzed neural measure for each trial of the experiment.

Appendix A.1. Implementing the Discrete State Representation

The discrete state representation assumes that the observed data were generated by two or more discrete latent states. In the main text, for example, we hypothesized that two latent states underlying task performance might correspond to an *on-task* state, where attention is directed to external stimuli such as an experimental task, or an *off-task* state, where attention is directed to internal stimuli such as mind wandering; these two states have been proposed in the popular perceptual decoupling hypothesis of mind wandering (Smallwood and Schooler, 2015). One could also hypothesize more than two discrete states; for example, Cheyne et al. (2009) hypothesized a three-state model of engagement/disengagement from task performance. However, for simplicity, we restricted the model recovery analyses in the main text and the outline here to the more prominent two-state case.

The partition of individual trials into the latent states can be derived from neural data in two main ways: using single measures, or multiple measures.

Appendix A.1.1. Identifying Discrete Latent States From A Single Neural Measure

Step 1: The first method begins with identification of a neural signal related to the latent states of interest. In the mind wandering literature, for example, activity of the default mode network (DMN) tends to increase during phases of off-task focus and decrease during phases of on-task focus (Andrews-Hanna et al., 2010; Christoff et al., 2009; Mason et al., 2007; Mittner et al., 2014; for meta-analysis, see Fox et al., 2015). In this case, the neural signal of interest could be a single-trial measure of DMN activity. The neural signal of interest can be simple in the sense that it involves a single measure (e.g., stimulus evoked pupil response, P3 ERP component over parietal cortex, or BOLD response in dorsolateral prefrontal cortex) or 'complex' in the sense that it involves an amalgamation of numerous measures (e.g., connectivity between various cortical regions); the key requirement is that a single value of the neural signal can be extracted for each trial (methods corresponding to multiple neural signals on each trial are discussed in the following subsection). The specifics for obtaining a single value on each trial might differ depending on the domain of study and the latent states of interest; it could be the value of the neural signal in a one second interval during the pre-stimulus period, immediately post-stimulus presentation, the full time course of a trial, or some other relevant interval.

Step 2: Once a single value of the neural signal is obtained for each trial, the individual trials are sorted in order of those with the lowest value of the neural signal (e.g., low DMN activity) through to those with the highest value of the neural signal (e.g., high DMN activity). Once sorted, the trials are split into separate groups. A simple option is to perform a median split of the DMN activity-sorted trials on the assumption that trials with lower DMN activity are more likely to have been generated by the on-task state and those with higher DMN activity are more likely to have been generated by the off-task state. A median split is a coarse approach and other methods can be used; for example, taking the lower 40% and upper 40% of trials, or using signs of bimodality in the distribution of the neural signal as an indicator of the appropriate cut point for the sorted trials. The key requirement is that the neural signal is used to split individual trials into at least two discrete groups of trials.

Step 3: Once the data have been split based on the neural signal, the cognitive model is fit to the discrete groups of trials. Critically, this fitting occurs *as if* the discrete groups were part of the experimental design. In the main text, for example, we assumed a single experimental condition with no explicit manipulation. When the data were split according to the neural signal, we essentially created a data set with two conditions that corresponded to the two latent states; we labeled these 'on task' and 'off task'. When fitting the model to the latent states one can estimate partially overlapping or distinct sets of model parameters for the discrete states. This is the same logic as fitting regular experimental manipulations: when difficulty is manipulated across conditions the conventional approach is to freely estimate a drift rate parameter for each condition while constraining other model parameters to a common value. In the discrete states case, one might hypothesize that drift rate differs across conditions but other parameters do not. This corresponds to the assumption that the latent states only differ in the efficiency of stimulus information processing.

The cognitive processes that might differ across latent states ought to be driven by theory. Ultimately, however, it comes down to a question of model selection; do processes A and B differ across latent states, or only process A? Such model comparison allows one to ask the question: if there are differences in cognitive processes between the latent states, what is the most likely difference? We argue that the final and most critical comparison is whether the simplest model of performance differences across the latent states is preferred to a model fit to data that is *not* split according to neural signal. The model recovery properties of this comparison were the primary focus of the main text.

Appendix A.1.2. Identifying Discrete Latent States From Multiple Neural Measures

The second method differs to the first in terms of the number of neural signals used to identify the latent cognitive states, and the complexity of the methods used to infer the latent state. The first method assumed that the neural signal collapsed to a single value for each trial. The second method attempts to combine multiple neural signals to infer the latent generating state on each trial. The general idea is that each neural signal might contain independent information about the latent state so simple methods of aggregation may lose discriminatory power. A more powerful form of aggregation is through supervised learning algorithms, though this places an additional requirement on data collection to obtain the 'labels' for to train a classification algorithm.

Step 1: The neural signals are extracted in a similar manner to Step 1 of *Appendix A.1.1.* However, here we assume there is a set of neural signals associated with the latent state of interest; the states might be on task and off task and the measures might be regional DMN activity, the task positive network, connectivity between the DMN and the task positive network, and stimulus evoked pupil diameter (cf. Mittner et al., 2014).

Step 2: The general approach outlined here was performed in Mittner et al. (2014). The aim is to collect the neural signals identified in Step 1 and behavioral data during regular task performance that also involves occasional behavioral indicators of the relevant latent states. In mind wandering research, for example, participants are typically periodically asked to report whether their focus was 'more on task' or 'more off task' in the preceding trial, though this is not asked on all trials. This method takes these self-report ratings as an indicator of the latent state – on task or off task – and uses them as labels to train a classification algorithm to 'learn' the distinct patterns of (the collection of) neural signals that discriminate on-task from off-task self-report ratings. Once the trained and validated, the algorithm probabilistically classifies *all* unlabeled trials to the on-task or off-task state, based on the correspondence between the neural signals on each unlabeled trial with the neural signal on the labeled trials.

Step 3: Once individual trials have been classified to the on-task or off-task states, the cognitive model is fit to the discrete groups of trials in the same manner as Step 3 of *Appendix A.1.1.* Typical classification algorithms used for Step 2 produce not only a latent state classification, but also a probability of correct assignment to the state (i.e., p_{on} and $p_{off} = 1 - p_{on}$). This uncertainty can be modeled in the liklihood function for each trial's data as a mixture of the likelihoods of the on-task and off-task states to account for noise in classification accuracy. Specifically, if the data from trial i are D_i and the likelihood of the set of on-task parameters given classification to the on-task state for D_i is $L(\theta_{on}|D_i, on - task)$, and similarly for off-task, then:

$$L(\theta|D_i) = p_{on,i}L(\theta_{on}|D_i, on - task) + p_{off,i}L(\theta_{off}|D_i, off - task)$$

Matthias: Are there any details that you think should be added to the machine learning outline? If so, can you please add some comments that detail what is missing.

Appendix A.2. Implementing the Continuous Dimension Representation

The continuous dimension representation assumes that the observed data were generated by a process that dynamically varies along a continuous latent dimension, relaxing the assumption that there are discrete latent states. In the context of mind wandering, for example, this approach assumes a trial will fall at some point along a continuum that spans from completely on-task focus through to completely off-task focus. The position along this latent continuum dynamically varies throughout the task.

The aim of this method is to regress a single-trial neural signal onto structured trial-by-trial variation in a model parameter. Here we outline and provide code to regress a single neural signal onto a single model parameter. However, the methods can be easily extended to regress multiple neural signals onto a single parameter (via multiple regression) or regress multiple neural signals onto multiple model parameters (via separate simple or multiple regressions for different model parameters).

Step 1: The neural signal is extracted in an identical manner to Step 1 of *Appendix A.1.1*. For the analyses described in the main text and outlined here, we assume that the neural signal is normalized to a Gaussian distribution with mean 0 and standard deviation 1. This permits examination of simple linear relationships for the mapping between the neural signal and the model parameter. Other forms of regression that do not assume simple linear mappings are possible but we do not explore those here.

Step 2: The hypothesized neural signal-model parameter mapping is formulated via simple linear regression. For example, in the main text we explored a covariate model that mapped a single-trial neural signal to single-trial drift rates (formulae 3 of the main text). Denote the normalized neural signal d, regression coefficient β, and drift rate v, then the simplest covariate model for drift rate on trial i is:

$$v_i \sim v + \beta \cdot d_i \tag{A.1}$$

(we also assumed across-trial variability in drift rate in the covariate model described in formulae 3 of the main text, which is omitted here for simplicity). This mapping assumes that the drift rate on trial i, v_i, has a mean component – the intercept, v, representing average performance in the condition/experiment – that is modulated on a trial-by-trial basis by the magnitude and valence of the neural signal on trial i, d_i, scaled by a regression coefficient, β, which is an index of effect size.

The neural signal can theoretically map to any parameter of the cognitive model of interest. When modulating single-trial parameter values it is important to ensure that the regression (A.1) does not allow any single-trial parameter estimates to move beyond feasible boundaries of the model (e.g., a single-trial value of the response threshold or non-decision time below 0). This can be instantiated with a 'check' in the parameter estimation routine that assigns very small likelihood to trials with infeasible single-trial parameter values, which results in low likelihood for the corresponding estimate of β. Alternatively the parameter can be transformed so that it is unbounded.

Step 3: Once the single-trial regression is parameterized, the cognitive model is fit to the behavioral data. In addition to other model parameters, this involves estimating parameters corresponding to the mean component and the regression coefficient of the linear regression (v and β in A.1, respectively). The neural signal (d_i) is provided with the data. Together, these three components allow estimation of a unique drift rate for each trial (v_i). The accompanying code provides explicit details how to compute this step.

In the context of single-trial regression there is an added interpretational benefit to using a Bayesian approach to parameter estimation: if the posterior distribution of β does not contain zero there is likely a significant effect of the neural signal on the model parameter. Other hypothesis tests are also possible using the posterior distribution, for example, estimation of the Savage-Dickey Bayes factor. Inference on β is less straightforward using conventional parameter estimation methods such as maximum likelihood estimation, though is still possible.

The extent to which the estimate of the β parameter differs to 0 gives an estimate of the significance of the neural signal on the model parameter, and hence cognitive process of interest. For example, if we regressed single-trial measures of (normalized) DMN activity onto drift rate and obtained an estimate of $\beta = -.2$, this indicates that for each unit increase in DMN activity there was a decrease of .2 in drift rate.

As in the discrete state analyses, the cognitive processes that might be dynamically modulated by a neural signal ought to be driven by theory. However, again, this comes down to a question of model selection; does the neural signal have a stronger single-trial effect on process A or B of the model? As before, we argue that the most important comparison is whether the most parsimonious single-trial regression model is preferred to a model fit to data that is *not* informed by a neural signal. The model recovery properties of this comparison were the primary focus of the main text.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychological Review 111, 1036–1060.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., Buckner, R. L., 2010. Functional–anatomic fractionation of the brain's default network. Neuron 65, 550–562.

Bastian, M., Sackur, J., 2013. Mind wandering at the fingertips: Automatic parsing of subjective states based on response time variability. Frontiers in Psychology 4, n/a. doi: 10.3389/fpsyg.2013.00573.

Borst, J. P., Anderson, J. R., 2015. The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models. NeuroImage 108, 60–73.

Brooks, S. P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–455.

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

Burnham, K. P., Anderson, D. R., 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research 33, 261–304.

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.

Cheyne, J. A., Solman, G. J., Carriere, J. S., Smilek, D., 2009. Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention–related errors. Cognition 111, 98–113.

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 of the United States of America 106, 8719–8724.

Craigmile, P. F., Peruggia, M., Van Zandt, T., Jun. 2010. Hierarchical Bayes Models for Response Time Data. Psychometrika 75 (4), 613–632.

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

Forstmann, B. U., Ratcliff, R., Wagenmakers, E.-J., 2016. Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. Annual Review of Psychology 67, 641–666.

Fox, K. C. R., Spreng, R. N., Ellamil, M., Andrews-Hanna, J. R., Christoff, K., 2015. The wandering brain: Meta–analysis of functional neuroimaging studies of mind–wandering and related spontaneous thought processes. NeuroImage 111, 611–621.

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., Badre, D., 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. The Journal of Neuroscience 35, 485–494.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Aki, V., Rubin, D. B., Sep. 2014. Bayesian Data Analysis, 3rd Edition. Texts in Statistical Science. CRC Press.

Gomez, P., Ratcliff, R., Perea, M., 2007. A model of the go/no–go task. Journal of Experimental Psychology: General 136, 389–413.

Hawkins, G. E., Hayes, B. K., Heit, E., 2016. A dynamic model of reasoning and memory. Journal of Experimental Psychology: General 145, 155–180.

Hawkins, G. E., Mittner, M., Boekel, W., Heathcote, A., Forstmann, B. U., 2015. Toward a model-based cognitive neuroscience of mind wandering. Neuroscience 310, 290–305.

Kass, R. E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Association 90, 377–395.

Killingsworth, M. A., Gilbert, D. T., 2010. A wandering mind is an unhappy mind. Science 330, 932–932.

Lee, M. D., 2011. How cognitive modeling can benefit from hierarchical Bayesian models. Journal of Mathematical Psychology 55, 1–7.

Logan, G. D., Van Zandt, T., Verbruggen, F., Wagenmakers, E.-J., 2014. On the ability to inhibit thought and action: General and special theories of an act of control. Psychological Review 121, 66–95.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., Macrae, C. N., 2007. Wandering minds: The default network and stimulus–independent thought. Science 315, 393–395.

McVay, J. C., Kane, M. J., 2010. Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). Psychological Bulletin 136, 188–207.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., Forstmann, B. U., 2014. When the brain takes a break: A model–based analysis of mind wandering. The Journal of Neuroscience 34, 16286–16295.

Nosofsky, R. M., Palmeri, T. J., 1997. An exemplar–based random walk model of speeded classification. Psychological Review 104, 266–300.

Nunez, M. D., Srinivasan, R., Vandekerckhove, J., 2015. Individual differences in attention influence perceptual decision making. Frontiers in Psychology 8, n/a. doi: 10.3389/fpsyg.2015.00018.

Nunez, M. D., Vandekerckhove, J., Srinivasan, R., in press. How attention influences perceptual decision making: Single-trial EEG correlates of drift–diffusion model parameters. Journal of Mathematical Psychology.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Raftery, A. E., 1995. Bayesian model selection in social research. Sociological Methodology, 111–196.

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

Ratcliff, R., McKoon, G., 2008. The diffusion decision model: Theory and data for two–choice decision tasks. Neural Computation 20, 873–922.

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

Ratcliff, R., Tuerlinckx, F., 2002. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychonomic Bulletin & Review 9, 438–481.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., Yiend, J., 1997. 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia 35, 747–758.

Schall, J. D., 2004. On building a bridge between brain and behavior. Annual Review of Psychology 55, 23–50.

Shiffrin, R. M., Steyvers, M., 1997. A model for recognition memory: REM-retrieving effectively from memory. Psychonomic Bulletin & Review 4, 145–166.

Singmann, H., Brown, S., Gretton, M., Heathcote, A., 2016. rtdists: Response time distributions. R package version 0.4-9.
URL `http://CRAN.R-project.org/package=rtdists`

Smallwood, J., Schooler, J. W., 2006. The restless mind. Psychological Bulletin 132, 946–958.

Smallwood, J., Schooler, J. W., 2015. The science of mind wandering: Empirically navigating the stream of consciousness. Annual Review of Psychology 66, 487–518.

Smilek, D., Carriere, J. S., Cheyne, J. A., 2010. Failures of sustained attention in life, lab, and brain: Ecological validity of the SART. Neuropsychologia 48, 2564–2570.

Smith, P. L., Ratcliff, R., 2004. The psychology and neurobiology of simple decisions. Trends in Neurosciences 27, 161–168.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64, 583–639.

Steingroever, H., Wetzels, R., Wagenmakers, E.-J., 2016. Bayes factors for reinforcement-learning models of the Iowa Gambing Task. Decision 3, 115–131.

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., Baddeley, A. D., 1995. Stimulus-independent thought depends on central executive resources. Memory & Cognition 23, 551–559.

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

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

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

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Vandekerckhove, J., Tuerlinckx, F., Lee, M. D., 2008. A Bayesian approach to diffusion models of decision–making. In: Sloutsky, V. M., Love, B. C., McRae, K. (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society. Cognitive Science Society, pp. 1429–1434.

Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. Psychonomic Bulletin & Review 11, 192–196.

Wagenmakers, E. J., Farrell, S., Ratcliff, R., 2004. Estimation and interpretation of 1/f α noise in human cognition. Psychonomic bulletin & review 11 (4), 579–615.

Watanabe, S., 2013. A widely applicable Bayesian information criterion. The Journal of Machine Learning Research.

Weissman, D. H., Roberts, K. C., Visscher, K. M., Woldorff, M. G., 2006. The neural bases of momentary lapses of 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, n/a. doi: 10.3389/fninf.2013.00014.