How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters

Michael D. Nunez^{a,c,*}, Joachim Vandekerckhove^{a,c,d}, Ramesh Srinivasan^{a,b,d}

^a*Department of Cognitive Sciences, University of California, Irvine, CA, USA*

^b*Department of Biomedical Engineering, University of California, Irvine, CA, USA*

^c*Department of Statistics, University of California, Irvine, CA, USA*

^d*Institute of Mathematical Behavioral Sciences, University of California, Irvine, CA, USA*

Abstract

Perceptual decision making can be accounted for by drift-diffusion models, a class of decision-making models that assume a stochastic accumulation of evidence on each trial. Fitting response time and accuracy to a drift-diffusion model produces evidence accumulation rate and non-decision time parameter estimates that reflect cognitive processes. Our goal is to elucidate the effect of attention on visual decision making. In this study, we show that measures of attention obtained from simultaneous EEG recordings can explain per-trial evidence accumulation rates and perceptual preprocessing times during a visual decision making task. Models assuming linear relationships between diffusion model parameters and EEG measures as external inputs were fit in a single step in a hierarchical Bayesian framework. The EEG measures were features of the evoked potential (EP) to the onset of a masking noise and the onset of a task-relevant signal stimulus. Single-trial evoked EEG responses, P200s to the onsets of visual noise and N200s to the onsets of visual signal, explain single-trial evidence accumulation and preprocessing times. Within-trial evidence accumulation variance was not found to be influenced by attention to the signal or noise. Single-trial measures of attention lead to better out-of-sample predictions of accuracy and correct reaction time distributions for individual subjects.

*Corresponding author

Email address: mdnunez1@uci.edu (Michael D. Nunez)

Keywords: Visual attention, Perceptual decision making, Diffusion Models, Neurocognitive modeling, Electroencephalography (EEG), Hierarchical Bayesian modeling

1. Introduction

There are many situations on the road when the driver of a vehicle must decide to stop or accelerate through an intersection by observing a traffic light. The presence of a green arrow for an adjacent lane (i.e. the distractor or “noise”) can be distracting for the driver whose light is red (i.e. the “signal”). The presence of the distractor affects the reaction time and choice of the driver. However the driver can suppress their attention to the green arrow and/or attend to the correct red light in their lane. The decision to stop or accelerate is an example of a perceptual decision. Perceptual decision making is the process of making quick decisions based on objects’ features observed with the senses. As shown in the stoplight example, attention is highly influential in the perceptual decision making process. When distracting objects exist in visual space, one must attend only to the relevant objects and actively ignore distracting objects. Each time an individual reaches a stop light, they will be more likely to make a safer decision if they suppress distracting visual input and enhance relevant visual input.

The goal of this study was to evaluate whether attention could predict different components of the decision making process on each trial of a visual discrimination experiment. We make use of high-density electroencephalographic (EEG) recordings from the human scalp to find single-trial evoked potentials (EPs) to the onset of visual signal and to the onset of a distractor (mask) to measure the deployment of attention to task-relevant features. We found that on each trial, modulations of the evoked potentials by attention were predictive of specific components of a drift-diffusion model of the decision making process.

1.1. Visual attention and decision making

Attention is beneficial for decision making because relevant features of the environment can be preferentially processed to enhance the quality of evidence. During visual tasks individuals may deploy different attention strategies such as: enhancing the signal, suppressing external noise (distractors), or suppressing internal noise (Lu and Doshier, 1998; Doshier and Lu,

2000). These strategies are thought to change based on the signal to noise ratio of the stimulus, such that individuals will enhance sensory gain to both signal and noise during periods of low noise and sharpen attention to only signal during periods of high noise (Lu and Doshier, 1998), although specific strategies have been shown to differ across subjects (Bridwell et al., 2013; Krishnan et al., 2013; Nunez et al., 2015). Multiple groups have proposed models of visual attention and decision making that yield diverse reaction time and choice distributions dependent upon attentional load (Spieler et al., 2000; Smith and Ratcliff, 2009). Attention can be deployed to the features and/or location of a stimulus, and attention can benefit decision making when the subject is cued to the location or features of the stimulus (Eriksen and Hoffman, 1972; Shaw and Shaw, 1977; Davis and Graham, 1981).

Event-related potentials (ERPs) are trial-averaged EEG responses to external stimuli. Visual ERPs (also labeled Visual Evoked Potentials; VEPs) have been shown to track visual attention to the onset of stimuli (Harter and Aine, 1984; Luck et al., 2000). That is, amplitudes of the peaks of the ERP waveform (i.e., ERP “components”) within certain millisecond-scale time windows are shown to be larger when subjects encounter task-relevant stimuli in the expected location in visual space. Two components of particular interest are the *N1* (or *N200*) and *P2* (or *P200*) components. The terms *N1* and *P2* refer the order of negative and positive peaks in the time series respectively, and the more general alternative names *N200* and *P200* refer to their approximate latencies in milliseconds. Changes in *N200* latencies have been shown to correlate with attentional load (Callaway and Halliday, 1982), and *N200* measures have even been used in Brain-Computer Interfaces (BCI) that make use of subjects’ attention to specific changing stimuli, such as letters in a BCI speller (Hong et al., 2009). Findings in these trial-averaged EEG (ERP) studies suggest that information is also available in single-trials of EEG that can be used to evaluate the relationship between attention and decision making. In this paper we will use the alternative names *P200* and *N200* because 1) the exact time windows of components vary across studies, 2) components in this study were both localized to around 200 milliseconds, and 3) components in this study were found on *single-trials* as opposed to in the trial-average.

1.2. Behavioral models of decision making

Drift-diffusion models are a widely-used class of models used to jointly predict subjects’ choices and reaction times (RT) during two-choice decision

69 making (Stone, 1960; Ratcliff, 1978; Ratcliff and McKoon, 2008). “Neural”
70 drift-diffusion models have also successfully incorporated functional magnetic
71 resonance imaging (fMRI) and EEG recordings into hierarchical models of
72 choice-RT (e.g. Mulder et al., 2014; Turner et al., 2015; Nunez et al., 2015). In
73 this study, we use a hierarchical form of the diffusion model (Vandekerckhove
74 et al., 2011), allowing variability between participants and across conditions,
75 to predict and describe single-trial reaction times and accuracy during a
76 visual decision making task. While other similar models of choice-RT have
77 successfully predicted behavior during visual decision making, such as the
78 simpler linear ballistic accumulator model (Brown and Heathcote, 2008) or
79 a more complicated drift-diffusion model that intrinsically accounts for trial-
80 to-trial variability in parameters within subjects (Ratcliff, 1978; Ratcliff and
81 McKoon, 2008), we have chosen a diffusion model that allows us to test
82 specific predictions from models of attention (i.e. Smith and Ratcliff, 2009;
83 Lu and Doshier, 1998) while being simple-enough to fit in reasonable time
84 periods given the hierarchical form.

85 In the drift-diffusion model it is assumed that subjects accumulate ev-
86 idence for one choice over another (or a correct versus incorrect response,
87 as in this study) in a random walk evidence accumulation process with an
88 infinitesimal time step. That is, evidence E_t accumulates following a Wiener
89 process (i.e. Brownian motion) with drift rate δ and instantaneous variance
90 ς^2 (Ross, 2014) such that

$$\frac{dE_t}{dt} \sim \mathcal{N}(\delta, \varsigma^2) \quad (1)$$

91 Thus the drift rate δ describes mean rate of evidence accumulation within a
92 trial and the diffusion coefficient ς influences the variance of evidence accu-
93 mulation within one trial, with the true variance of the current evidence at
94 any particular time t being $\varsigma^2 t$. A graphical representation of the diffusion
95 model is provided in the middle panel of **Figure 1**.

96 A few other parameters describe a diffusion model. The boundary separa-
97 tion α is equal to the amount of relative evidence required to make choice
98 A over choice B (or make a correct decision over an incorrect decision), and
99 the boundary separation has been shown to be manipulated by speed vs.
100 accuracy strategy trade-offs (Voss et al., 2004). The parameter that encodes
101 the starting position of evidence β reflects the bias towards one choice or
102 another (equal to .5 when modeling correct versus incorrect choices as in
103 this study). The non-decision time τ is equal to the amount of time within

the reaction time of each trial that is not dedicated to the decision making process. Typically non-decision time is assumed to be equal to the sum of preprocessing time before the evidence accumulation process and motor response time after a decision has been reached. The relative contribution of these two non-decision times is not identifiable from behavior alone and therefore is rarely explicitly modeled.

All three of the parameters related to evidence accumulation are not identifiable with behavioral data alone (i.e. drift rate δ , the diffusion coefficient ς , and the boundary separation a). Only two of the three parameters can be assumed to vary across subjects and trials (e.g. multiplying ς by two and dividing both a and δ by two would result in the same fit of choice-RT) (Ratcliff and McKoon, 2008; Wabersich and Vandekerckhove, 2014). Previous studies have typically chosen to fix the diffusion coefficient ς to 1 or 0.1 (Vandekerckhove et al., 2011; Wabersich and Vandekerckhove, 2014). However due to the predictions made by Doshier and Lu (2000), in that internal noise is suppressed by attention to the signal, we choose to leave ς to vary. The boundary separation α was fixed at 1 for all trials and subjects. Our primary analysis focused on the trial-to-trial variability in the evidence accumulation process due to fluctuations in attention from trial-to-trial within individuals. Although trial-to-trial speed-accuracy trade-offs can be experimentally introduced to find neural correlates of the boundary separation (e.g. van Maanen et al., 2011) or may exist due to per-trial performance feedback (Dutilh et al., 2012), we have no reason to believe that the boundary separation will vary considerably from trial-to-trial within a subject due to changes in *attention*. Moreover, all subjects were given the same accuracy instruction to maintain similar accuracy vs. speed trade-offs across subjects.

1.3. Single-trial EEG measures of attention

EEG correlates of attention and decision making have been found using classification methods. One group has shown that the amplitude of certain EEG components in the time domain track type and duration of two-alternative forced choice responses and then showed that these components' amplitudes tracked evidence accumulation rate (Philiastides and Sajda, 2006; Ratcliff et al., 2009). However the EEG components in these studies were found by finding the maximum predictors of the behavioral data and thus had no a priori interpretation. Another group has found that that single-trial amplitude in a few frequency bands, especially the 4-9 Hz theta band, predicts evidence accumulation rates (van Vugt et al., 2012). However these

141 oscillations were found using canonical correlation analysis (CCA; Calhoun
 142 et al., 2001), a data driven algorithm that found any EEG channel mixtures
 143 that contained correlations with the drift diffusion model parameters. While
 144 the results were confirmed using cross validation, the set of EEG identified
 145 by this method also did not have an a priori explanation. These studies point
 146 us in directions of exploration and perform well at prediction, but we have
 147 little information as to whether the EEG information reflected attention, the
 148 decision process itself, or some other correlate of evidence accumulation. In
 149 this study, we introduce a simple procedure that is informed by ERPs known
 150 to be related to attention, and we make use of single-trial ERP estimates to
 151 model behavior on single trials.

152 1.4. Hypothesized attention effects

153 An integrated model of visual attention, visual short term memory, and
 154 perceptual decision making by Smith and Ratcliff (2009) predicts that atten-
 155 tion operates on the encoding of the stimulus, and that enhanced encoding
 156 increases drift rate during the decision making process. Furthermore, the
 157 model predicts that visual encoding time (i.e. visual preprocessing) will be
 158 reduced by attention which is reflected in the non-decision time parameter.
 159 However, this model of visual attention only considers the detection of a
 160 stimulus in an otherwise blank field—that is, a field with no visual noise.
 161 Thus, it does not have predictions for the distinct processes of noise suppres-
 162 sion and signal enhancement, as in the Perceptual Template Model (Lu and
 163 Doshier, 1998). Signal enhancement during the evidence accumulation pro-
 164 cess is predicted to reduce the diffusion coefficient ς because the mechanism
 165 by which signal enhancement takes place, according to the Perceptual Tem-
 166 plate Model, is additive *internal noise* reduction¹ (Doshier and Lu, 2000);
 167 this mechanism is predicted to be most effective in low noise conditions since
 168 decreasing internal noise will lead to better processing of *both* the visual sig-
 169 nal and external visual noise. External noise suppression, on the other hand,
 170 is expected to reflect the encoding of the stimulus by manipulation of a per-
 171 ceptual template, increasing the average rate of evidence accumulation δ by

¹“In stimulus enhancement, attention increases the gain on the stimulus, which is formally equivalent to reducing internal additive noise. This can improve performance only in low external noise stimuli, since external noise is the limiting factor in high external noise stimuli.” (Doshier and Lu, 2000, p. 1272)

improving the overall quality of evidence on a trial. The Perceptual Template Model predicts this mechanism is most effective in high noise conditions.

In a previous study we showed that *individual differences* in noise suppression predicts individual differences in evidence accumulation rates and non-decision times (Nunez et al., 2015). We also showed that differences across individuals in signal enhancement predict individual differences in non-decision times and evidence accumulation variance (i.e. the diffusion coefficient), which we assume tracks internal noise in the subject. The findings of signal enhancement effects on evidence accumulation variance and noise suppression effects on evidence accumulation rate seem to correspond closely to predictions made by the Perceptual Template Model. However the Perceptual Template Model does not make explicit predictions about attention effects on non-decision times. The previous study did not explore how trial-to-trial variation in attention affected trial-to-trial cognitive differences. Individual differences in attention could be found that are not detected to be changing within a subject, and/or trial-to-trial variability in attention could occur that does not change across individuals. In this study, we show that within-subject, trial-to-trial variability in attention to *both* noise and signal predict variability in drift rate and non-decision times, corresponding closely to predictions made by the model of Smith and Ratcliff (2009) that predicts speeded encoding time and increased evidence accumulation rate due to enhanced attention. The two studies together suggest that within-trial evidence accumulation variances ς varied across individuals, but we did not find evidence that this measured varied within individuals due to changes in trial-to-trial attention.

2. Methods

2.1. Experimental stimulus: Bar field orientation task

Reported in a previous study, behavioral and EEG data were collected from a simple two-alternative forced choice task to test *individual differences* in attention during visual decision making (Nunez et al., 2015). Here, we reanalyze these data to explore *per-trial* attention effects on the decision making process. Subjects were instructed on each trial to differentiate the mean rotation of a field of small bars that were either oriented at 45 deg or 135 deg from horizontal on average. Two representative frames of the display and the time course of a trial are provided in **Figure 2**. The circular field of small bars was embedded in a square field of visual noise that was changing

208 at 8 Hz. The bar field was flickering at 15 Hz. These frequencies were chosen
209 to evoke steady-state visual evoked potentials (SSVEPs), stimulus frequency-
210 tagged EEG responses that were useful in the previous study but will not
211 be used in this study. Stimuli were built and displayed using the MATLAB
212 Psychophysics toolbox (Psychtoolbox-2; www.psychtoolbox.org).

213 Subjects viewed each trial of the experimental stimulus on a monitor in
214 a dark room. Subjects sat 57 cm away from the monitor. The entire circular
215 field of small oriented bars was 9.5 cm in diameter, corresponding to a visual
216 angle of 9.5° . Within each trial subjects first observed a black cross for 750 ms
217 in the center of the screen, on which they were instructed to maintain fixation
218 throughout the trial. Subjects then observed visual noise for 750 ms. This
219 time period of the stimulus will henceforth be referred to as *cue interval*, with
220 the onset EEG response at the beginning of this interval being the response
221 to the *noise* (or “distractor”) stimulus. Subjects then observed the circular
222 field of small oriented bars overlaid on the square field of visual noise for 1000
223 to 2000 ms and responded during this interval. Subjects were instructed to
224 respond as accurately as possible while providing a response during every
225 trial. Because evidence required to make a decision only appeared during
226 this time frame, the decision process was assumed to take place during this
227 interval. This interval is referred to as the *response interval*, and attentional
228 onset EEG measures during this time period are referred to as responses to
229 the *signal* stimulus. After the response interval the fixation cross was again
230 shown for 250 ms to alert the subjects that the trial was over and to collect
231 any delayed responses.

232 Three levels of variance of bar rotation and three levels of noise luminance
233 were used to modulate the task difficulty. However only the noise luminance
234 manipulation is relevant for the analysis presented here. Average luminance
235 of the noise was 50% and the luminance of the bars was 15%. In the low
236 noise condition, that luminance was drawn randomly at each pixel from a
237 uniform distribution of 35% to 65% luminance. In the medium and high noise
238 conditions, noise luminance was drawn randomly at each pixel from a uniform
239 distribution of 27.5% to 72.5% and 20% to 80% luminance respectively. Each
240 subject experienced 180 trials from each noise condition, interleaved, for a
241 total of 540 trials split evenly over 6 blocks. The total duration of the visual
242 experiment for each subject was approximately one hour and 15 minutes
243 including elective breaks between blocks. More details of the experiment can
244 be found in our previous publication (Nunez et al., 2015).

245 Behavioral and EEG data were collected concurrently from 17 subjects.

Subjects performed accurately during the task. The across-subject mean, standard deviation, and median of accuracy were $90.1\% \pm 5.8\%$, $\tilde{y} = 91.6\%$, while the across-subject mean, standard deviation, and median of average reaction time were 678 ± 106 , $\tilde{t} = 670$ ms. Individual differences in behavior existed across subjects with the most accurate subject answering 98.3% of trials correctly and the least accurate subject answering 78.5% of trials correctly. Two different subjects were the fastest and slowest with mean RTs of 502 ms and 866 ms respectively.

2.2. Single-trial EEG predictors

Electroencephalograms (EEG) were recorded using Electrical Geodesics, Inc.'s high density 128-channel Geodesic Sensor Net and Advanced Neuro Technology's amplifier. Electrical activity from the scalp was recorded at a sampling rate of 1024 samples per second with an online average reference using Advanced Neuro Technology software. The EEG data was then imported into MATLAB for offline analysis. Linear trends were removed from the EEG data, and the data were band pass filtered to a 1 to 50 Hz window using a high pass Butterworth filter (1 Hz pass band with a 1 dB ripple and a 0.25 Hz stop band with 10 dB attenuation) and a low pass Butterworth filter (50 Hz pass band with 1 dB ripple and a 60 Hz stop band with 10 dB attenuation).

EEG *artifact* is broadly defined as data collected within EEG recordings that does not originate from the brain. Electrical artifact can be biological (e.g. from the muscles-EMG or from the arteries-EKG) or non-biological (e.g. temporary electrode dislocations, DC shifts, or 60 Hz line noise). Contribution of muscle and electrical artifact was reduced in recordings by using an extended Infomax Independent Component Analysis algorithm (ICA; Makeig et al., 1996; Lee et al., 1999). ICA algorithms are used to find linear mixtures of EEG data that relate to specific artifact. Components that are indicative of artifact typically have high spatial frequency scalp topographies, high temporal frequency or a $1/f$ frequency falloff, and are present either in only a few trials or intermittently throughout the recording. These properties are not shared by electrical signals from the brain as recorded on the scalp (Nunez and Srinivasan, 2006). Using these metrics, components manually deemed to reflect artifact were projected into EEG space and subtracted from the raw data. Components deemed to be a mixture of artifact and brain activity were kept. More information about using ICA to reduce the contribution of artifact can be found in Jung et al. (2000).

283 Event-related potential (ERP) components have been shown to index at-
284 tention (Callaway and Halliday, 1982; Harter and Aine, 1984; Luck et al.,
285 2000), in particular the P200 and N200 latencies and amplitudes, and these
286 values were used as independent measures of attention in the following analy-
287 ses. Event-related potentials (ERPs) are EEG responses that are time-locked
288 to a stimulus onset and are typically estimated by aligning and averaging
289 EEG responses across trials. They usually cannot be directly measured on
290 each trial from single electrodes. Raw EEG signals could be used as a single-
291 trial measures but typically have very low signal-to-noise ratios (SNR) for
292 task-specific brain responses. Since the goal of this analysis was to explore
293 single-trial effects of attention on visual decision making, a single-trial esti-
294 mate of the ERP was developed.

295 Because the signal-to-noise ratio (SNR) in ongoing EEG increases when
296 adjacent electrodes of relevant activity are summed (Parra et al., 2005), we
297 anticipated that the SNR of the response to the visual stimulus would be
298 boosted on individual trials by summing over the mixture of channels that
299 best described the average visual response. Traditional ERPs at each chan-
300 nel (represented by a matrix of size $T \times C$ where T is the length of a trial
301 in milliseconds and C is the number of EEG channels) were calculated sep-
302 arately for each subject. One ERP was calculated for the response to the
303 visual signal and another was found for the response to the visual noise by
304 averaging a random set of two-thirds of the trials across all conditions for
305 each subject in each window. This random set of trials was the same set
306 used as the training set for cross validation, to be discussed later. The test
307 sets of trials were not used to calculate the traditional ERPs.

308 Singular value decomposition (SVD; analogous to principal component
309 analysis) of the *trial averages* were then used to find linear mixtures of chan-
310 nels that explained the largest amount of the variance in the ERP data (i.e.
311 the first right-singular vectors $\mathbf{v}$, explaining a percentage of variance from
312 39.4% to 91.9% and 45.0% to 93.2% across subjects in the cue and response
313 intervals respectively). The first right-singular vectors were then used as
314 weights to mix the raw EEG data into a brain response biased toward the
315 maximum response to the visual stimuli, yielding one time course of the EEG
316 per trial for both the cue and response intervals. A visual representation of
317 the simple procedure for a single trial is provided in **Figure 3**. The raw data
318 matrix $\mathbf{E}$ of dimension $N \times C$ was multiplied by the first right singular vector
319 $\mathbf{v}$ (a $C \times 1$ vector of channel weights) to obtain a $N \times 1$ vector $\mathbf{E}\mathbf{v} = \mathbf{e}$, which
320 could then be split up into epochs of length $T \times 1$ representing the response

321 to the stimulus on each trial. Note that the voltage amplitudes of the ERP
 322 measures calculated based on this method will differ from traditional single
 323 electrode ERP amplitudes since the single-trial estimates are a weighted sum
 324 of potentials over all electrodes.

325 Not only did this method boost the SNR of the EEG measures, but this
 326 method also reduced EEG measures of size $T \times C$ on each trial to one latent
 327 variable that varies in time of size $T \times 1$. Thus the correlation of the EEG as
 328 inputs to the model was drastically reduced and the interpretability of model
 329 parameters was increased compared to analyses with highly correlated model
 330 inputs. The weight vector $\mathbf{v}$ for each subject in both the cue and response
 331 intervals also yields a scalp map when the values of the weights are interpo-
 332 lated between electrodes. Channel weights calculated using SVD on subject's
 333 ERPs to the noise onset (during the cue interval) are shown in topographic
 334 plots for each subject in **Figure 7**. Channel weights calculated using SVD on
 335 subject's ERPs to the signal onset (during the response interval) are shown
 336 in topographic plots for each subject in **Figure 6**. While raw EEG on single
 337 trials from single electrodes may have large enough SNRs to be informative
 338 for our analysis, we would not obtain an idea of the locus of activation or the
 339 pattern of activation over the scalp.

340 We focused our analysis on the windows 150 to 275 ms post stimulus-
 341 onset in the cue and response intervals. These windows were found to con-
 342 tain P200 and N200 ERP components. On each trial, we measured the peak
 343 positive and negative amplitudes, and the latency at which these peaks were
 344 observed. We used these 8 single-trial measures to predict single-trial dif-
 345 fusion model parameters. However in this paper we will focus only on the
 346 results of models with 4 single-trial measures: the amplitude and latency of
 347 the peak positive deflection (P200) during the *cue interval* and the ampli-
 348 tude and latency of the peak negative deflection (N200) during the *response*
 349 *interval*, because very weak evidence, if any, was found for the effects of the
 350 other attention measures on diffusion model parameters in models with all 8
 351 single-trial measures. It should be noted that single-trial measures of EEG
 352 spectral responses at SSVEP frequencies (see Nunez et al., 2015) were briefly
 353 explored but future methods must be developed to increase signal-to-noise
 354 ratios of SSVEP measures on single-trials.

355 2.3. Hierarchical Bayesian models

356 Hierarchical models of visual decision making were assumed and placed
 357 into a Bayesian framework. Bayesian methods yield a number of benefits

358 compared to other inferential techniques such as traditional maximum likeli-
359 hood methods. Rather than point estimates of parameters, Bayesian methods
360 provide entire distributions of the unknown parameters. Bayesian methods
361 also allow us to perform the model fitting procedure in a single step, main-
362 taining all uncertainty about each parameter through each hierarchical level
363 of the model.

364 One downside of Bayesian methods is that creating sampling algorithms
365 to find posterior distributions of Bayesian hierarchical models can be time
366 consuming and cumbersome. However Just Another Gibbs Sampler (JAGS;
367 Plummer et al., 2003) is a program that uses multiple sampling techniques to
368 find estimates of hierarchical models, only requiring the form of the model,
369 data, and initial values as input from the user. In order to find posterior dis-
370 tributions, we have used JAGS with an extension that adds a diffusion model
371 distribution (without intrinsic trial-to-trial variability) as one of the distri-
372 butions to be sampled from (Wabersich and Vandekerckhove, 2014). Similar
373 software packages to fit hierarchical diffusion models have been developed in-
374 dependently in other programming languages such as Python (Wiecki et al.,
375 2013).

376 In order to evaluate the benefit to prediction of adding EEG measures to
377 hierarchical diffusion models, three different models were compared. Model
378 3 assumed that evidence accumulation rates, evidence accumulation vari-
379 ances, and non-decision times were each equal to a linear combination of
380 EEG measures on each trial. Because we found no effect of the observed
381 single-trial EEG measures on single-trial evidence accumulation variances,
382 we also fit Model 2, where single-trial evidence accumulation rates and non-
383 decision times were influence by EEG, but single-trial evidence variances
384 were not. Model 1 did not assume any EEG contribution to any parameters.
385 This model assumed that parameters not varying with EEG would change
386 based on subject and condition, drawn from a condition level distribution.
387 Graphical representations of the hierarchical Bayesian models are provided
388 in **Figure 4** following the convention of Lee and Wagenmakers (2014).

389 For Model 1 (**Figure 4a**), prior distributions were kept mostly uninforma-
390 tive (i.e. parameters of interest had prior distributions with large variances)
391 so that the analyses would be data-driven. The prior distributions of pa-
392 rameters for each subject j and condition k free from EEG influence had the
393 following prior and hyperprior structure

394

$$\begin{aligned}
& (\delta_{jk} \mid \mu_{(\delta)k}, \sigma_{(\delta)}) \sim \mathcal{N}(\mu_{(\delta)k}, \sigma_{(\delta)}^2), \quad \mu_{(\delta)k} \sim \mathcal{N}(1.5, 4^2) \in (-9, 9), \quad \sigma_{(\delta)} \sim \Gamma(5, 0.20) \\
& (\tau_{jk} \mid \mu_{(\tau)k}, \sigma_{(\tau)}) \sim \mathcal{N}(\mu_{(\tau)k}, \sigma_{(\tau)}^2), \quad \mu_{(\tau)k} \sim \mathcal{N}(0.3, 1^2) \in (0, 3) \quad , \quad \sigma_{(\tau)} \sim \Gamma(5, 0.05) \\
& (\varsigma_{jk} \mid \mu_{(\varsigma)k}, \sigma_{(\varsigma)}) \sim \mathcal{N}(\mu_{(\varsigma)k}, \sigma_{(\varsigma)}^2), \quad \mu_{(\varsigma)k} \sim \mathcal{N}(0.6, 2^2) \in (0, 4) \quad , \quad \sigma_{(\varsigma)} \sim \Gamma(5, 0.05)
\end{aligned}$$

Where the normal distributions $\mathcal{N}$ are parameterized with mean and variance respectively and the gamma distributions Γ are parameterized with shape and scale parameters respectively.

In Models 2 (**Figure 4b**) and 3 (**Figure 4c**), to ensure noninterference by the prior distributions, uninformative priors were given for both the effects γ_{jk} of EEG on the parameters of interest and the linear intercepts η_{jk} . Note that the effect of EEG γ_{jk} is a vector with one element per EEG regressor and each effect of EEG is assumed to be statistically independent from the others. If a drift-diffusion model parameter was assumed to be equal to a linear combination of EEG inputs then the following two lines replaced the priors of the respective parameter above.

$$\begin{aligned}
& (\eta_{jk} \mid \mu_{(\eta)k}, \sigma_{(\eta)}) \sim \mathcal{N}(\mu_{(\eta)k}, \sigma_{(\eta)}^2) \quad , \quad \mu_{(\eta)k} \sim \mathcal{N}(0, 100^2) \quad , \quad \sigma_{(\eta)} \sim \Gamma(5, 5) \\
& (\gamma_{jk} \mid \boldsymbol{\mu}_{(\gamma)k}, \sigma_{(\gamma)}) \sim \mathcal{MVN}(\boldsymbol{\mu}_{(\gamma)k}, \sigma_{(\gamma)}^2 I), \quad \boldsymbol{\mu}_{(\gamma)k} \sim \mathcal{MVN}(\mathbf{0}, 100^2 I), \quad \sigma_{(\gamma)} \sim \Gamma(5, 5)
\end{aligned}$$

In Models 2 and 3, the parameter on each trial was assumed to be equal to a simple linear combination of the vector of single-trial EEG inputs $\mathbf{x}_{ijk}$ on that trial i with η_{jk} and γ_{jk} as the intercept and slopes respectively:

$$\begin{aligned}
& \delta_{ijk} = \eta_{(\delta)jk} + \boldsymbol{\gamma}_{(\delta)jk}^\top \mathbf{x}_{ijk} \\
& \tau_{ijk} = \eta_{(\tau)jk} + \boldsymbol{\gamma}_{(\tau)jk}^\top \mathbf{x}_{ijk} \\
& \varsigma_{ijk} = \eta_{(\varsigma)jk} + \boldsymbol{\gamma}_{(\varsigma)jk}^\top \mathbf{x}_{ijk}
\end{aligned}$$

Where the first two equations refer to the structure of Model 2 and all three equations refer to the structure of Model 3. Note that the $p * 1$ vector of effects γ_{jk} of EEG on each parameter could include the intercept term η_{jk} to create a $(p + 1) * 1$ vector of effects γ_{jk}^* (and the EEG vector $\mathbf{x}_{ijk}$ would include a value of 1 to be multiplied by the intercept term). We use this notation in **Figure 4** for simplicity.

Because not all trials are believed to actually contain a decision-making process (i.e. the subject quickly presses a random button during a certain percentage of trials reflecting a “fast guess”), reaction times below a certain threshold were removed from analysis and cross-validation. Cutoff reaction times were found for each subject by using an exponential moving average of accuracy after sorting by reaction time (Vandekerckhove and Tuerlinckx,

2007). The rejected reaction times were all below 511 ms with a mean cutoff of 410 ms across subjects. This resulted in an average rejection rate of 1.4% of trials across subjects with a maximum of 6.3% of trials rejected for one subject and a minimum of 0.7% of trials rejected for 11 of the 17 subjects.

Each model was fit using JAGS with six Markov Chain Monte Carlo (MCMC) chains run in parallel (Tange, 2011) of 52,000 samples each with 2,000 burn-in samples and a thinning parameter of 10 resulting in 5,000 posterior samples in each chain. The posterior samples from each chain were combined to form one posterior sample of 30,000 samples for each parameter. All three models converged as judged by $\hat{R}$ being less than 1.02 for all parameters in each model. $\hat{R}$ is a statistic used to assess convergence of MCMC algorithms (Gelman and Rubin, 1992).

Posterior distributions were found for each free parameter in the three models. Credible intervals of the found posterior distributions were then calculated to summarize the findings of each model. EEG regressor effects were deemed to have weak evidence if the 95% credible interval between the 2.5th and 97.5th percentiles of the subject mean parameter $\mu_{(\gamma)j}$ was non-overlapping zero. Effects were deemed to have strong evidence if the 99% credible interval between the 0.5th and 99.5th percentiles was non-overlapping zero.

2.4. Cross-validation

All trials from all subjects were used during initial exploration of the data. However once it was decided that the signal onset response was a candidate predictor of drift rate, cross-validation was performed using a training and test set of trials. Out-of-sample performance for both *known* and *unknown* subjects were found by randomly assigning two-thirds of the trials from each subject in a random sample of *subjects* (i.e. 13 of 17 subjects were chosen at random) as the training set and one-third of the trials from the 13/17 subjects and all trials from the remaining 4/17 subjects as the test set. After fitting the model with the training set, posterior predictive distributions of the accuracy-RT data were found for each subject. Posterior predictive distributions were calculated by drawing from the subject-level posteriors of the *known* subjects and by drawing from the condition-level posteriors of the *unknown* subjects. The posterior predictive distributions were then compared to the sample distribution of the test set.

In some recent papers, evaluation of models' prediction ability has been left to the readers with the aid of posterior predictive coverage plots (e.g. see

466 figures in Supplementary Materials). Here we formally evaluate the similarity
 467 of the posterior predictive distributions to the test samples via a “propor-
 468 tion of variance explained” calculation. Specifically, we calculated R^2_{pred} of
 469 subjects’ accuracy and correct reaction time 25th percentiles, medians, and
 470 75th percentiles across subjects. R^2_{pred} is a measure of percentage variance in
 471 a statistic T (e.g. accuracy, correct-RT median, etc.) explained by in-sample
 472 or out-of-sample prediction. In this paper, R^2_{pred} is defined as the percentage
 473 of total *between-subject variance* of a statistic T explained by out-of-sample
 474 or in-sample prediction. It is a function of the mean squared error of predic-
 475 tion (MSEP) and the sample variance of the statistic T based on a sample
 476 size of $J = 13$ or $J = 4$ subjects for known and unknown subject calcula-
 477 tions respectively. This measure also allows comparisons across studies with
 478 similar prediction goals. Mathematically, R^2_{pred} is defined as

$$R^2_{\text{pred}} = 1 - \frac{\sum_{j=1}^J (T_j - T_{(\text{pred})j})^2 / (J - 1)}{\sum_{j=1}^J (T_j - \bar{T})^2 / (J - 1)} = 1 - \frac{\text{MSEP}_T}{\widehat{\text{Var}[T]}} \quad (2)$$

479 **3. Results**

480 The single-trial EEG measures “regressed” on diffusion model parameters
 481 were the peak positive and negative amplitudes and latencies (correspond-
 482 ing to P200 and N200 peaks respectively) in the 150 to 275 ms windows
 483 post noise-onset in the cue interval and post signal-onset in the response in-
 484 terval. However the magnitude and latency of the peak negative deflection
 485 (N200) in response to the noise stimulus and the magnitude and latency of
 486 the peak positive deflection (P200) in response to the signal stimulus were
 487 not informative (i.e. most condition-level effect posteriors of these measures
 488 overlapped zero significantly in models with all P200 and N200 measures in-
 489 cluded as regressors). For simplicity we only discuss results of models with
 490 P200 measures following the noise stimulus in the cue interval and N200 mea-
 491 sures following the signal stimulus in the response interval. Example single
 492 trial amplitudes of these P200 and N200 peaks for Subject 12 are shown in
 493 **Figure 5**.

494 Since no effects of explored measures were found on within-trial evidence
 495 accumulation variance in Model 3 (i.e. posterior distributions of $\gamma_{(\varsigma)jk}$ were
 496 centered near zero), the only EEG effects that will be discussed are those
 497 on evidence accumulation rate and non-decision time from a fit of Model
 498 2. The posterior distributions of EEG effects on evidence accumulation rate

499 $\gamma_{(\delta)jk}$ and non-decision time $\gamma_{(\tau)jk}$ did not differ significantly from Model 2
 500 to Model 3. Moreover, Model 2 produced better out-of-sample prediction
 501 than Model 3 for new subjects. A graphical example of the effects found
 502 with Model 2 in two representative trials are given in **Figure 1**.

503 *3.1. Intercept terms of evidence accumulation rate and non-decision time*

504 The intercept term of each variable gives the value of each variable not
 505 explained by a linear relationship to N200 and P200 amplitudes and latencies.
 506 That is, the intercept gives the value of each parameter that remains constant
 507 from trial to trial, with the between-trial variability of the parameter being
 508 influenced by the changing trial-to-trial EEG measures. Model 2's posterior
 509 medians of the condition level evidence accumulation rate intercepts $\mu_{(\eta_\delta)j}$
 510 and non-decision time intercepts $\mu_{(\eta_\tau)j}$ are reported. In low noise conditions,
 511 evidence accumulation rate intercepts were 1.46 evidence units per second
 512 (i.e. if there was no behavioral effect of EEG on each trial and no variance in
 513 the evidence accumulation process, it would take the average subject 343 ms
 514 to accumulate evidence since a decision is reached when $\alpha = 1$ evidence unit
 515 is accumulated and subjects start the evidence accumulation process with .5
 516 evidence units). In medium and high noise conditions, evidence accumulation
 517 rate intercepts were 1.30 and 0.86 evidence units per second respectively.
 518 Non-decision time intercepts were 340 ms in low noise conditions, 425 ms in
 519 medium noise conditions, and 440 ms in high noise conditions.

520 To understand the degree of influence of EEG on model parameters, ap-
 521 proximate condition level evidence accumulation rates and non-decision times
 522 were calculated and then compared to the intercept of the respective param-
 523 eter. Taking the mean peak positive and peak negative amplitudes and la-
 524 tencies across all subjects and trials in each noise condition and multiplying
 525 by the median posterior of the effects, it was found that evidence accumulate
 526 rate in low noise was 1.90 evidence units per second, 1.65 evidence units per
 527 second in medium noise, and 1.35 evidence units per second in high noise. It
 528 was also found that non-decision time was 393, 400, and 425 ms in the low,
 529 medium, and high noise conditions. The intercepts of non-decision time thus
 530 described approximately 86%, 94%, and 96% of the true condition means
 531 in low, medium, high noise conditions respectively. However, the intercepts
 532 of the drift rates only described approximately 77%, 79%, and 63% of the
 533 true condition means in low, medium, and high noise conditions respectively.
 534 While this gives the reader an idea of the strength of the influence of single-
 535 trial EEG measures on the parameters, better evaluations of the degree of

536 effects are presented below.

537 3.2. *Effects of attention on non-decision time in low-noise conditions*

538 Strong evidence was found to suggest that in low noise conditions single-
539 trial non-decision times τ_{ijk} are positively linearly related to delays in the
540 EEG response to the visual signal as measured by the latency of the negative
541 peak (N200) following stimulus onset. A probability greater than 99% of the
542 condition-level effect being greater than zero in all subjects was found. This
543 relationship to an EEG signature 150-275 ms post stimulus onset suggests an
544 effect on preprocessing time rather than motor-response time. By exploring
545 the posterior distribution of the mean effect across-participants $\mu_{(\gamma_\tau)_j}$, it is
546 inferred that non-decision time increases 12 ms (the posterior median) when
547 there is a 40 ms increase in the latency of the single-trial negative peak (where
548 40 ms was the standard deviation across all trials and subjects) in the low
549 noise condition, with a 99% credible interval of 3 to 21 ms. **Figure 6** shows
550 the per-subject effects of signal N200 latency on non-decision time in the
551 low noise condition. No evidence was found to suggest that the signal N200
552 latency affected non-decision time in medium nor high noise conditions. 95%
553 credible intervals for these increases in the subject mean non-decision time for
554 40 ms increased N200 delays were -9 to 4 ms and -8 to 3 ms respectively.
555 No evidence was found to suggest that attentional delay to the noise, the
556 noise P200 latency, affected non-decision time.

557 Weak evidence was found to suggest that magnitude of the response to
558 the stimulus affects non-decision time in the low noise condition. The pos-
559 terior median suggests that a $26.83 \mu V$ (the standard deviation) decrease in
560 magnitude of the negative peak (i.e. moves the negative peak towards zero)
561 leads to a 11 ms increase in non-decision time. The 95% credible interval of
562 this effect of N200 signal magnitude on non-decision time was a 2 to 21 ms.
563 No evidence was found to suggest that the magnitude in the medium and
564 high noise conditions affected non-decision time.

565 3.3. *Effects of attention on evidence accumulation*

566 Evidence was found to suggest that per-trial response to the visual signal
567 (measured by the negative peak, N200, amplitude) is positively correlated
568 with per-trial evidence accumulation rates δ_{ijk} in each condition. In the low
569 noise condition, $\mu_{(\gamma_\delta)_j}$, which describes the across-subject mean of the effect
570 of negative peak on drift rate, was found to have a 95% credible interval of .02
571 to .34 evidence units per second increase (where it takes $\alpha = 1$ evidence unit

to make a decision) and a posterior median of .17 evidence units per second increase for each magnitude increase (i.e. away from zero) of 26.83 μV , the standard deviation of the negative peak. Given the same magnitude increase, the posterior median of the effects in the medium and high noise conditions were .13 and .14 evidence units per second respectively with 95% credible intervals $-.01$ to $.28$ and 0 to $.28$ respectively.

Strong evidence was found to suggest that the magnitude of the positive peak of the response to the *visual noise* during the cue interval affected the future evidence accumulation rate in the medium noise and possibly high noise conditions. The median of the posterior distribution of the condition-level effect was .20 evidence units per second when there was a 27.67 μV increase, the standard deviation of the peak magnitude. A 99% credible interval of this effect was .04 to .32 evidence units per second. **Figure 7** shows the effects of the noise P200 amplitudes on specific subjects' single-trial drift rates in the medium noise condition. The probability of there being an effect of this P200 amplitude during the cue interval in the *high* noise condition was 94.6% (i.e. the amount of the posterior density of the condition-level effect above zero). The median of the posterior distribution of this effect was .09 evidence units per second with a 95% credible interval of $-.02$ to $.22$ evidence units per second when there is a 27.67 μV increase in a high noise trial.

3.4. Cross-validation

In-sample and out-of-sample posterior predictive coverage plots of correct-RT distributions for each condition and subject are provided in the Supplementary Materials. All three models perform well at predicting correct-RT distributions and overall accuracy of training data (i.e. *in-sample* prediction; see **Table 3** and Discussion section). However by cross-validation we found that the addition of single-trial EEG measures of attentional onset improved *out-of-sample* prediction of accuracy and correct reaction time distributions of *known* subjects (i.e. those subjects who had 2/3 of their trials used in the training set). R^2_{pred} indicates the percentage of variance explained by prediction in the given statistic across subjects. **Table 1** contains R^2_{pred} values for accuracy as well as summary statistics of correct-RT distributions of known subjects. Prediction was improved when using Model 3 for these subjects with at least 77.3% of variance in correct-RT medians being explained by out-of-sample prediction, but Model 2 performed almost as well in comparison to Model 1, the model without single-trial EEG inputs. Model 2 was able

609 to predict at least 76.3% of the variance in correct-RT medians while Model
610 1 was able to predict at least 74.5% of the variance in correct-RT medians.
611 In the low noise condition, Model 2 did not improve upon Model 1’s expla-
612 nation of variance in subject-level accuracy but better predicted accuracy in
613 the other two conditions.

614 Larger gains in out-of-sample prediction were found for *unknown* subjects
615 (i.e. those subjects who were not used in the training set). These improve-
616 ments were particularly pronounced in the low noise condition. Model 2
617 outperformed Model 3, which outperformed Model 1 in turn, as shown in
618 **Table 2**. From these results it is clear that Model 2 was the best model
619 for out-of-sample prediction overall, especially for new subjects. In the low
620 noise condition, Model 2 was able to explain 22.1% of between-subject vari-
621 ance in correct-RT 25th percentiles in the low-noise condition while Model 1
622 was not able to predict any between-subject variance in this measure. While
623 the included single-trial EEG measures in these type of models do not per-
624 form as well as new subject prediction when *subject-average* EEG measures
625 of attention are included (Nunez et al., 2015), single-trial EEG does improve
626 prediction. The improvements in R^2_{pred} across models suggest that it is possi-
627 ble that similar models with more single-trial measures of EEG could explain
628 new subjects’ accuracy and correct-RT distributions.

629 3.5. P200 and N200 localizations

630 Both the subject-average P200 and N200 components were localized in
631 time and space. All *single-trial* P200 and N200 amplitudes and latencies
632 (i.e. single-trial peak positive and negative amplitudes 150 to 275 ms in
633 the noise and response interval respectively) were averaged across trials for
634 each subject. Localization in time and on the brain are based on these
635 across-trial averages. The subject mean and standard deviation of the P200
636 latency during the cue interval was 220 ± 12 ms while the mean and standard
637 deviation the of N200 latency during the response interval was 217 ± 6 ms.
638 Although these latencies differ slightly from traditional P2 and N1 findings
639 (Luck et al., 2000), when viewing the event-related waveforms over all trials
640 it is clear that the P2 and N1 are influenced by the single-trial measures.
641 As an example, every single-trial evoked response of Subject 12 to the noise
642 and signal are shown in **Figure 5**, sorted by peak P200 amplitude in the cue
643 interval and sorted by peak N200 amplitude in the response interval. The
644 P200 and N200 latencies of this subject correspond to traditional P2 and N1
645 components.

646 While EEG localization is an inexact process that is unsolvable with-
647 out additional assumptions, the surface Laplacian has been shown to match
648 closely to simulated cortical activity using forward models (i.e the mapping
649 of cortical activity to scalp potentials) and have shown consistent results
650 when used with real EEG data (Nunez and Srinivasan, 2006). Unlike 3D so-
651 lutions, projections to the surface of the cortex (more accurately, the dura)
652 are theoretically solvable, and have been used with success in past studies
653 (see Nunez et al., 1994, for an example).

654 In this study we have found surface spline-Laplacians (Nunez and Pil-
655 green, 1991) on the realistic MNI average scalp (Deng et al., 2012) of the
656 mean positive peak during the cue interval (the P200) and the mean nega-
657 tive peak during the response interval (the N200) by averaging over trials and
658 subjects. The surface Laplacians were then projected onto one subject’s cor-
659 tical surface using Tikhonov (L2) regularization and a Finite Element (FE;
660 Pommier and Renard, 2005) forward model to the MNI 151 average head,
661 maintaining similar distributions of activity of the surface Laplacians on the
662 cortical surface. The subject’s brain was then labeled using the Destrieux
663 cortical atlas (Fischl et al., 2004). Cortical topographic maps of both peaks
664 are given in **Figure 8**. Because we expect the majority of the Laplacian
665 to originate from superficial gyri (Nunez and Srinivasan, 2006), we have lo-
666 calized both the positive and negative peaks only to maximally active gyri.
667 This localization suggested that both the P200 and N200 were in the follow-
668 ing extrastriate and parietal cortical locations: right and left middle occipital
669 gyri, right and left superior parietal gyri, right and left angular gyri, the left
670 occipital inferior gyrus, the right occipital superior gyrus, and the left tem-
671 poral superior gyrus. Although we should note that the exact localization
672 must have some errors due to between-subject variance in cortex and head
673 shape and between-subject variance in tissue properties.

674 Brain regions found using this cortical-Laplacian method point to activ-
675 ity in early dorsal and ventral pathway regions associated with visual atten-
676 tion (Desimone and Duncan, 1995; Corbetta and Shulman, 2002; Buschman
677 and Miller, 2007) and decision making (Mulder et al., 2014). Corroborating
678 our findings, White et al. (2014) found that blood-oxygen-level dependent
679 (BOLD) activity in the right temporal superior gyrus, right angular gyrus,
680 and areas in the right lateral occipital cortex (e.g. the right middle occipital
681 gyrus) correlated with non-decision time during a simple visual and auditory
682 decision making task. It was hypothesized that this activity was due to motor
683 preparation time instead of visual preprocessing time (White et al., 2014);

684 however the time-scale of BOLD signals does not provide additional knowl-
685 edge to separate visual preprocessing time from motor preparation time. In-
686 formed by EEG, BOLD signals associated with evidence accumulation rates
687 have been previously localized to right and left superior temporal gyri and
688 lateral occipital cortical areas, thought to correspond to early bottom-up and
689 late top-down decision making processes respectively during a visual face/car
690 discrimination task (Philiastides and Sajda, 2007). The right and left mid-
691 dle occipital gyri have also previously been shown to contribute to evidence
692 accumulation rates during a random dot motion task (Turner et al., 2015).

693 4. Discussion

694 4.1. *Attention influences perceptual decision making on each trial*

695 The results of this study suggest that fluctuations in attention to a vi-
696 sual signal accounts for some of the trial-to-trial variability in the brain’s
697 speed of evidence accumulation on each trial in each condition. There is also
698 evidence to suggest that increased response to the competing visual noise in-
699 creases the brain’s speed of evidence accumulation, but only in medium and
700 high noise conditions. Although a simple explanation of this effect would
701 be differences in trial-to-trial arousal, we note that the effect only occurs in
702 medium-noise and high-noise conditions. We have previously found evidence
703 that noise suppression during the cue interval predicts enhanced drift rates
704 based on the subject-average SSVEP responses in these data (Nunez et al.,
705 2015). Thus, our effect may reflect the attention the subject places on the
706 cue, which determines whether to engage mechanisms of noise suppression,
707 but we could not directly assess this possibility as new methods must be
708 developed to measure SSVEPs on single trials. This would reflect a hypoth-
709 esis based on the Perceptual Template Model that predicts that subjects will
710 suppress attention to visual distractors during tasks of high visual noise (Lu
711 and Doshier, 1998).

712 We assume that the effect of N200 latency on non-decision time during
713 the response interval is on preprocessing time instead of motor response time.
714 There was no a priori reason to believe that an attention effect that takes
715 place 150-275 ms after stimulus onset would affect the speed of efferent sig-
716 nals to the muscle, given that the response times were at least 500 ms. These
717 findings lead us to conclude that the effect of attention in response to the
718 signal in low noise conditions is to reduce preprocessing delay time. This

appears consistent with predictions of signal enhancement in low-noise conditions in the Perceptual Template Model. However, we note there is not a perfect equality between N200 latency and preprocessing time (the condition and subject-level coefficient posteriors are not centered on 1). And the identification of both preprocessing time and motor response time is not possible with a drift-diffusion model fit without additional assumptions or external inputs such as EEG.

Reaction time (RT) and choice behavior during visual decision making tasks are well characterized by models that assume a continuous stochastic accumulation of evidence. And many observations of increasing spike-rates of single neuron action potentials lend support to this stochastic theory of evidence accumulation on a neural level (Shadlen and Newsome, 1996, 2001). Recently some macroscopic recordings of the cortex have shown that increasing EEG potentials ramping up to P300 amplitudes are correlates of the stochastic accumulation of evidence (O’Connell et al., 2012; Philiastides et al., 2014; Twomey et al., 2015). It has been hypothesized that this EEG data reflects the evidence accumulation process itself (or a mixture of this process with other decision-making correlates) and not a correlated measure such as top-down attention. This hypothesis leads to the natural prediction that single-trial drift rates are explained by single-trial P300 slopes. However within a small region of the cortex, neurons will have diverse firing patterns during the decision making process, only some of which are observed to have increasing spiking-rate behavior indicative of stochastic evidence accumulation (Meister et al., 2013). The properties of volume conduction through the cortex, skull, and skin only allow for synchronous post-synaptic potentials to be observed at the scalp (Nunez and Srinivasan, 2006; Buzsaki, 2006). Therefore an increasing spike-rate as observed on the single-neuron level is not likely to be observed as an increasing waveform in EEG recorded from the scalp. We also would expect the evidence accumulation process to terminate before the response time since a portion of the response time must be dedicated to the motor response after the decision is made. This may not be the case for the ramping P300 waveform on single-trials even though it is predictive of model parameters (Philiastides et al., 2014). If the stochastic evidence accumulation process was truly reflected in the ramping of EEG, a testable prediction would be that the variance around the mean rate of the P300 ramp on each trial would be linearly related to the diffusion coefficient ς , in addition to single-trial P300 slopes being linearly related to the drift rate δ . There are many EEG measures thought to be related to *attention* such as

757 event related potential (ERP) components, power in certain frequency bands,
758 and steady-state visual evoked potential (SSVEP) responses. It is likely that
759 EEG measures that share similar properties with stochastic evidence accu-
760 mulation processes are in fact due to these correlates of attention or other
761 forms of cortical processing that can influence the decision making process.

762 4.2. External predictors allow for trial level estimation of diffusion model 763 parameters

764 The Wiener distribution (i.e. the diffusion model) used in this study does
765 not incorporate trial-to-trial variability in drift rates *within the probability*
766 *density function* as assumed by Ratcliff (1978). Instead we assume that
767 each trial’s drift rate is *exactly equal to* a linear function of EEG data and
768 use an evidence accumulation likelihood function that does not assume drift
769 rates vary trial-to-trial by any other means. Per-trial non-decision times and
770 diffusion coefficients were also assumed to be *exactly equal to* linear functions
771 of EEG data.

772 Per trial estimates of diffusion model parameters cannot be obtained with-
773 out imposing constraints or including external inputs. In this study, we have
774 shown that the single-trial P200 and N200 attention measures can be used
775 to discover per trial estimates of all three free parameters, non-decision time,
776 drift rate, and the diffusion coefficient. Non-decision time, the drift rate, and
777 the *evidence boundary* could also be modeled as per trial estimates of external
778 inputs, as would be useful in other speeded reaction time tasks where exter-
779 nal per-trial physiological measurements are available. Other possible per-
780 trial external inputs that could be used include: magnetoencephalographic
781 (MEG) measures, functional magnetic resonance imaging (fMRI) measures,
782 physiological measurements such as galvanic skin response (GSR), and near
783 infrared spectroscopy (NIRS), where each modality may have multiple ex-
784 ternal inputs (e.g. multiple linear EEG regressors of single-trial parameters,
785 as in this study). The more external inputs correlated with single-trial pa-
786 rameters included in the model, the better the single-trial estimate of the
787 parameters will be. This will allow decision model researchers to better ex-
788 plore the efficacy of the diffusion model by comparing single-trial estimates
789 of parameters.

790 4.3. Behavior prediction and BCI applications

791 We have observed that single-trial measures of EEG in a hierarchical
792 Bayesian approach to decision-making modeling improves overall accuracy

793 and correct-RT distribution prediction for subjects with observed behav-
794 ior. This paradigm also lead to significant improvement in overall accuracy
795 and correct-RT distribution prediction for those subjects whose behavior
796 was *missing*. That is, we have shown that a new subject’s accuracy and
797 correct-RT distributions can be predicted when only their single-trial EEG
798 is collected, given that other subjects’ EEG and behavior has been analyzed.
799 If the goal of a future project is solely prediction (and no explanation of
800 the cognitive or neural process is desired, as was in this paper), a whole
801 host of single-trial EEG measures could be included in a hierarchical model
802 of decision making, using perhaps a simpler model of decision making such
803 as the linear ballistic accumulator model to ease analysis (e.g. Forstmann
804 et al., 2008; Ho et al., 2009; van Maanen et al., 2011; Rodriguez et al., 2015)
805 or a more complicated model to improve prediction. The set of single-trial
806 measures could include: ERP-like components as we discussed in this paper,
807 measures of evoked amplitudes in certain frequency bands, and measures of
808 steady-state visual evoked potentials.

809 However, as observed in **Table 3**, we have not shown prediction of re-
810 action times for when subjects committed errors because 1) few errors were
811 committed by the subjects in the presented data and 2) the type of model
812 we used has been shown to explain incorrect-RT distributions well only when
813 intrinsic trial-to-trial variability in evidence accumulation rates (as opposed
814 to *extrinsic* due to neural regressors, for which it has not been shown) is in-
815 cluded in the likelihood function (Ratcliff, 1978; Ratcliff and McKoon, 2008).
816 In future work we plan to compare the predictive ability of more compli-
817 cated models containing *both* intrinsic trial-to-trial variability and extrinsic
818 trial-to-trial variability in evidence accumulation rate due to external neural
819 measures.

820 The prediction ability of the presented models for accuracy and correct-
821 RT distributions may have implications for Brain-Computer Interface (BCI)
822 frameworks, especially in paradigms which attempt to enhance a participant’s
823 visual attention to particular task to improve reaction time. To maximize
824 prediction, every EEG attention measure should be included, and the pre-
825 dictors that offer the best out-of-sample prediction within initial K-fold val-
826 idation sets should be included. After collection of behavior and EEG from
827 a few participants and a hierarchical Bayesian analysis of the data, later
828 participants’ single-trial preprocessing times or evidence accumulation rates
829 could be predicted using only single-trial EEG measures. Conceivably this
830 would allow for trial-by-trial intervention in order to enhance a participant’s

831 attention during the task, perhaps using neural feedback (e.g. via direct
832 current stimulation or transcranial magnetic stimulation) during those trials
833 in which a participant is predicted to be slow in their response because of a
834 small evidence accumulation rate (e.g. when an N200 magnitude is small)
835 or a slow preprocessing time (e.g. when an N200 latency is long). Although
836 whether the participant could use such feedback in time to affect reaction
837 time and accuracy remains to be tested.

838 4.4. *Neurocognitive models*

839 The term “neurocognitive” has been used to describe the recent trend
840 of combining mathematical behavioral models and observations of brain be-
841 havior to explain and predict perceptual decision making (Palmeri et al.,
842 2014). The usefulness of combining behavioral models and neural dynam-
843 ics has been motivated on theoretical grounds. Behavioral models suggest
844 links between subject behavior and cognition while laboratory observation
845 and neuroimaging can suggest links between neural dynamics and cognition.
846 The combination of these methods then provides a predictive chain of neural
847 dynamics, cognition, and behavior. Another obvious benefit is the inference
848 gain when predicting missing data. That is, we will be able to better predict
849 behavior when brain activity is available. This is especially true when us-
850 ing hierarchical Bayesian models as they maintain uncertainty in estimates
851 through different levels of the analysis (Vandekerckhove et al., 2011; Turner
852 et al., 2013). While there are a variety of methods using cognitive models
853 to find cognitive correlates in the brain dynamics (see Turner et al., 2016,
854 for a review of these methods), some studies do not further constrain the
855 cognitive models by informing those models with known neural links to spe-
856 cific cognitive processes. In this paper, and our previous study of individual
857 differences (Nunez et al., 2015), we demonstrate another important use of
858 neural data in cognitive models. Independent *neural* measures of cognitive
859 processes, such as attention, can be used to better understand how cognition
860 influences the mechanisms of behavior, furthering explanation and prediction
861 of the cognitive process.

862 **Data and code sharing**

863 Pre-calculated EEG measures, raw behavioral data, MATLAB stimulus
864 code, JAGS code, and an example single-trial EEG R script are available

865 upon request and in the following repository (as of February 2016) if their
866 use is properly cited.
867 <https://github.com/mdnunez/mcntoolbox/>

868 **Acknowledgments**

869 Special thanks to Josh Tromberg and Aishwarya Gosai for help with EEG
870 collection and artifact removal. Cort Horton is well appreciated for creating
871 useful MATLAB functions for EEG visualization and artifact removal. Siyi
872 Deng and Sam Thorpe are thanked for their contributions to the FEM solu-
873 tions and anatomical fMRI image generation. The anonymous reviewers are
874 appreciated for their constructive suggestions and comments.

875 **Funding**

876 MDN and RS were supported by NIH grant 2R01MH68004. JV was
877 supported by NSF grants #1230118 and #1534472 from the Methods, Mea-
878 surements, and Statistics panel and grant #48192 from the John Templeton
879 Foundation.

880 **Conflict of interest statement**

881 The authors declare that the research was conducted in the absence of any
882 commercial or financial relationships that could be construed as a potential
883 conflict of interest.

884 **References**

- 885 Z.-L. Lu, B. A. Doshier, External noise distinguishes attention mechanisms,
886 Vision research 38 (1998) 1183–1198.
- 887 B. A. Doshier, Z.-L. Lu, Noise exclusion in spatial attention, Psychological
888 Science 11 (2000) 139–146.
- 889 D. A. Bridwell, E. A. Hecker, J. T. Serences, R. Srinivasan, Individual
890 differences in attention strategies during detection, fine discrimination, and
891 coarse discrimination, Journal of neurophysiology 110 (2013) 784–794.

- 892 L. Krishnan, A. Kang, G. Sperling, R. Srinivasan, Neural strategies for selec-
893 tive attention distinguish fast-action video game players, *Brain topography*
894 26 (2013) 83–97.
- 895 M. D. Nunez, R. Srinivasan, J. Vandekerckhove, Individual differences in
896 attention influence perceptual decision making, *Frontiers in Psychology* 6
897 (2015).
- 898 D. H. Spieler, D. A. Balota, M. E. Faust, Levels of selective attention revealed
899 through analyses of response time distributions., *Journal of Experimental*
900 *Psychology: Human Perception and Performance* 26 (2000) 506.
- 901 P. L. Smith, R. Ratcliff, An integrated theory of attention and decision
902 making in visual signal detection., *Psychological review* 116 (2009) 283.
- 903 C. W. Eriksen, J. E. Hoffman, Temporal and spatial characteristics of selec-
904 tive encoding from visual displays, *Perception & psychophysics* 12 (1972)
905 201–204.
- 906 M. L. Shaw, P. Shaw, Optimal allocation of cognitive resources to spatial
907 locations., *Journal of Experimental Psychology: Human Perception and*
908 *Performance* 3 (1977) 201.
- 909 E. T. Davis, N. Graham, Spatial frequency uncertainty effects in the detec-
910 tion of sinusoidal gratings, *Vision research* 21 (1981) 705–712.
- 911 M. R. Harter, C. J. Aine, Brain mechanisms of visual selective attention,
912 *Varieties of attention* (1984) 293–321.
- 913 S. J. Luck, G. F. Woodman, E. K. Vogel, Event-related potential studies of
914 attention, *Trends in cognitive sciences* 4 (2000) 432–440.
- 915 E. Callaway, R. Halliday, The effect of attentional effort on visual evoked
916 potential n1 latency, *Psychiatry Research* 7 (1982) 299–308.
- 917 B. Hong, F. Guo, T. Liu, X. Gao, S. Gao, N200-speller using motion-onset
918 visual response, *Clinical neurophysiology* 120 (2009) 1658–1666.
- 919 M. Stone, Models for choice-reaction time, *Psychometrika* 25 (1960) 251–260.
- 920 R. Ratcliff, A theory of memory retrieval, *Psychological review* 85 (1978)
921 59.

- 922 R. Ratcliff, G. McKoon, The diffusion decision model: theory and data for
923 two-choice decision tasks, *Neural computation* 20 (2008) 873–922.
- 924 M. Mulder, L. Van Maanen, B. Forstmann, Perceptual decision
925 neurosciences—a model-based review, *Neuroscience* 277 (2014) 872–884.
- 926 B. M. Turner, L. van Maanen, B. U. Forstmann, Informing cognitive ab-
927 stractions through neuroimaging: The neural drift diffusion model., *Psy-
928 chological review* 122 (2015) 312.
- 929 J. Vandekerckhove, F. Tuerlinckx, M. D. Lee, Hierarchical diffusion models
930 for two-choice response times, *Psychological Methods* 16 (2011) 44.
- 931 S. D. Brown, A. Heathcote, The simplest complete model of choice response
932 time: linear ballistic accumulation, *Cognitive psychology* 57 (2008) 153–
933 178.
- 934 S. M. Ross, *Introduction to probability models*, Academic press, 2014.
- 935 A. Voss, K. Rothermund, J. Voss, Interpreting the parameters of the diffusion
936 model: An empirical validation, *Memory & Cognition* 32 (2004) 1206–
937 1220.
- 938 D. Wabersich, J. Vandekerckhove, Extending jags: a tutorial on adding
939 custom distributions to jags (with a diffusion model example), *Behavior
940 research methods* 46 (2014) 15–28.
- 941 L. van Maanen, S. D. Brown, T. Eichele, E.-J. Wagenmakers, T. Ho, J. Ser-
942 ences, B. U. Forstmann, Neural correlates of trial-to-trial fluctuations in
943 response caution, *The Journal of Neuroscience* 31 (2011) 17488–17495.
- 944 G. Dutilh, J. Vandekerckhove, B. U. Forstmann, E. Keuleers, M. Brysbaert,
945 E.-J. Wagenmakers, Testing theories of post-error slowing, *Attention,
946 Perception, & Psychophysics* 74 (2012) 454–465.
- 947 M. G. Philiastides, P. Sajda, Temporal characterization of the neural corre-
948 lates of perceptual decision making in the human brain, *Cerebral cortex*
949 16 (2006) 509–518.
- 950 R. Ratcliff, M. G. Philiastides, P. Sajda, Quality of evidence for perceptual
951 decision making is indexed by trial-to-trial variability of the eeg, *Proceed-
952 ings of the National Academy of Sciences* 106 (2009) 6539–6544.

- 953 M. K. van Vugt, P. Simen, L. E. Nystrom, P. Holmes, J. D. Cohen, Eeg
954 oscillations reveal neural correlates of evidence accumulation, *Frontiers in*
955 *neuroscience* 6 (2012).
- 956 V. Calhoun, T. Adali, G. Pearlson, J. Pekar, A method for making group
957 inferences from functional mri data using independent component analysis,
958 *Human brain mapping* 14 (2001) 140–151.
- 959 B. A. Doshier, Z.-L. Lu, Mechanisms of perceptual attention in precuing of
960 location, *Vision research* 40 (2000) 1269–1292.
- 961 S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, et al., Independent
962 component analysis of electroencephalographic data, *Advances in neural*
963 *information processing systems* (1996) 145–151.
- 964 T.-W. Lee, M. Girolami, T. J. Sejnowski, Independent component analysis
965 using an extended infomax algorithm for mixed subgaussian and super-
966 gaussian sources, *Neural computation* 11 (1999) 417–441.
- 967 P. L. Nunez, R. Srinivasan, *Electric fields of the brain: the neurophysics of*
968 *EEG*, Oxford university press, 2006.
- 969 T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. Mckeown, V. Iragui,
970 T. J. Sejnowski, Removing electroencephalographic artifacts by blind
971 source separation, *Psychophysiology* 37 (2000) 163–178.
- 972 L. C. Parra, C. D. Spence, A. D. Gerson, P. Sajda, Recipes for the linear
973 analysis of eeg, *Neuroimage* 28 (2005) 326–341.
- 974 M. Plummer, et al., Jags: A program for analysis of bayesian graphical
975 models using gibbs sampling, in: *Proceedings of the 3rd international*
976 *workshop on distributed statistical computing*, volume 124, Technische
977 *Universit at Wien*, p. 125.
- 978 T. V. Wiecki, I. Sofer, M. J. Frank, Hddm: hierarchical bayesian estimation
979 of the drift-diffusion model in python, *Frontiers in neuroinformatics* 7
980 (2013).
- 981 M. D. Lee, E.-J. Wagenmakers, *Bayesian cognitive modeling: A practical*
982 *course*, Cambridge University Press, 2014.

983 J. Vandekerckhove, F. Tuerlinckx, Fitting the ratcliff diffusion model to
984 experimental data, *Psychonomic bulletin & review* 14 (2007) 1011–1026.

985 O. Tange, Gnu parallel - the command-line power tool, *The USENIX Mag-*
986 *azine* 36 (2011) 42–47.

987 A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple
988 sequences, *Statistical science* (1992) 457–472.

989 P. Nunez, R. Silberstein, P. Cadusch, R. Wijesinghe, A. Westdorp, R. Sriniv-
990 *asan*, A theoretical and experimental study of high resolution eeg based
991 on surface laplacians and cortical imaging, *Electroencephalography and*
992 *clinical neurophysiology* 90 (1994) 40–57.

993 P. L. Nunez, K. L. Pilgreen, The spline-laplacian in clinical neurophysi-
994 ology: a method to improve eeg spatial resolution., *Journal of Clinical*
995 *Neurophysiology* 8 (1991) 397–413.

996 S. Deng, W. Winter, S. Thorpe, R. Srinivasan, Improved surface laplacian
997 estimates of cortical potential using realistic models of head geometry,
998 *Biomedical Engineering, IEEE Transactions on* 59 (2012) 2979–2985.

999 J. Pommier, Y. Renard, Getfem++, an open source generic c++ library for
1000 finite element methods, 2005.

1001 B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H.
1002 *Salat*, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, et al., Automat-
1003 ically parcellating the human cerebral cortex, *Cerebral cortex* 14 (2004)
1004 11–22.

1005 R. Desimone, J. Duncan, Neural mechanisms of selective visual attention,
1006 *Annual review of neuroscience* 18 (1995) 193–222.

1007 M. Corbetta, G. L. Shulman, Control of goal-directed and stimulus-driven
1008 attention in the brain, *Nature reviews neuroscience* 3 (2002) 201–215.

1009 T. J. Buschman, E. K. Miller, Top-down versus bottom-up control of atten-
1010 tion in the prefrontal and posterior parietal cortices, *science* 315 (2007)
1011 1860–1862.

1012 C. N. White, E. Congdon, J. A. Mumford, K. H. Karlsgodt, F. W. Sabb,
1013 N. B. Freimer, E. D. London, T. D. Cannon, R. M. Bilder, R. A. Poldrack,
1014 Decomposing decision components in the stop-signal task: a model-based
1015 approach to individual differences in inhibitory control, *Journal of cogni-*
1016 *tive neuroscience* 26 (2014) 1601–1614.

1017 M. G. Philiastides, P. Sajda, Eeg-informed fmri reveals spatiotemporal char-
1018 acteristics of perceptual decision making, *The Journal of Neuroscience* 27
1019 (2007) 13082–13091.

1020 M. N. Shadlen, W. T. Newsome, Motion perception: seeing and deciding,
1021 *Proceedings of the National Academy of Sciences* 93 (1996) 628–633.

1022 M. N. Shadlen, W. T. Newsome, Neural basis of a perceptual decision in the
1023 parietal cortex (area lip) of the rhesus monkey, *Journal of neurophysiology*
1024 86 (2001) 1916–1936.

1025 R. G. O’Connell, P. M. Dockree, S. P. Kelly, A supramodal accumulation-
1026 to-bound signal that determines perceptual decisions in humans, *Nature*
1027 *neuroscience* 15 (2012) 1729–1735.

1028 M. G. Philiastides, H. R. Heekeren, P. Sajda, Human scalp potentials reflect a
1029 mixture of decision-related signals during perceptual choices, *The Journal*
1030 *of Neuroscience* 34 (2014) 16877–16889.

1031 D. M. Twomey, P. R. Murphy, S. P. Kelly, R. G. O’Connell, The classic
1032 p300 encodes a build-to-threshold decision variable, *European Journal of*
1033 *Neuroscience* (2015).

1034 M. L. Meister, J. A. Hennig, A. C. Huk, Signal multiplexing and single-
1035 neuron computations in lateral intraparietal area during decision-making,
1036 *The Journal of Neuroscience* 33 (2013) 2254–2267.

1037 G. Buzsaki, *Rhythms of the Brain*, Oxford University Press, 2006.

1038 B. U. Forstmann, G. Dutilh, S. Brown, J. Neumann, D. Y. Von Cramon,
1039 K. R. Ridderinkhof, E.-J. Wagenmakers, Striatum and pre-sma facili-
1040 tate decision-making under time pressure, *Proceedings of the National*
1041 *Academy of Sciences* 105 (2008) 17538–17542.

- 1042 T. C. Ho, S. Brown, J. T. Serences, Domain general mechanisms of perceptual
1043 decision making in human cortex, *The Journal of Neuroscience* 29 (2009)
1044 8675–8687.
- 1045 C. A. Rodriguez, B. M. Turner, T. Van Zandt, S. M. McClure, The neural
1046 basis of value accumulation in intertemporal choice, *European Journal of*
1047 *Neuroscience* 42 (2015) 2179–2189.
- 1048 T. J. Palmeri, J. D. Schall, G. D. Logan, Neurocognitive modeling of percep-
1049 tual decision making, *Oxford handbook of computational and mathemat-*
1050 *ical psychology*, 2014.
- 1051 B. M. Turner, B. U. Forstmann, E.-J. Wagenmakers, S. D. Brown, P. B.
1052 Sederberg, M. Steyvers, A bayesian framework for simultaneously model-
1053 ing neural and behavioral data, *NeuroImage* 72 (2013) 193–206.
- 1054 B. M. Turner, B. U. Forstmann, B. C. Love, T. J. Palmeri, L. Van Maanen,
1055 Approaches to analysis in model-based cognitive neuroscience, *Journal of*
1056 *Mathematical Psychology* (2016).
- 1057 S. D. Brown, R. Ratcliff, P. L. Smith, Evaluating methods for approximating
1058 stochastic differential equations, *Journal of mathematical psychology* 50
1059 (2006) 402–410.

1060 **Figures**

Prediction of new data from <i>known</i> subjects				
		Model 1	Model 2	Model 3
		Comparison	EEG- δ, τ	EEG- δ, τ, ς
Low	25 th $\mathbf{t}_1$ Percentile	81.6%	85.2%	85.6%
	$\mathbf{t}_1$ Median	74.5%	76.3%	77.3%
	75 th $\mathbf{t}_1$ Percentile	60.2%	63.3%	63.8%
	Accuracy	24.5%	20.8%	27.7%
Medium	25 th $\mathbf{t}_1$ Percentile	84.1%	85.1%	86.1%
	$\mathbf{t}_1$ Median	86.8%	87.6%	88.9%
	75 th $\mathbf{t}_1$ Percentile	63.1%	68.2%	69.3%
	Accuracy	58.1%	63.5%	63.5%
High	25 th $\mathbf{t}_1$ Percentile	73.0%	76.4%	76.6%
	$\mathbf{t}_1$ Median	77.4%	76.8%	77.8%
	75 th $\mathbf{t}_1$ Percentile	71.0%	74.2%	74.2%
	Accuracy	46.3%	48.9%	51.3%

Table 1: Percentage of across-subject variance explained by out-of-sample prediction (R^2_{pred}) for accuracy and summary statistics of correct-RT distributions of those subjects' that were included in the training set. 13 of the subjects' data were split into 2/3 training and 1/3 test sets. Posterior predictive distributions that predicted test set behavior were generated for 13 of the subjects by drawing from posterior distributions generated by the training set. In the Low, Medium, and High noise conditions, the 25th, 50th (the median), and 75th percentiles and means were predicted reasonably well by the model without single-trial measures of EEG, Model 1. However including single-trial measures of EEG improved prediction of correct-RT distributions, especially in the Low noise condition, with Model 3 (which assumes evidence accumulation rate, non-decision time, and evidence accumulation variance vary with EEG per-trial) only slightly outperforming Model 2 (which assumes evidence accumulation rate and non-decision time vary with EEG per-trial).

Figure 1: Two trials of Subject 10's SVD weighted EEG (Top and Bottom with bounds -85 to $85 \mu V$) and representations of this subject's evidence accumulation process on 6 low noise trials (Middle). Evidence for a correct response in one example trial (denoted by the red line) first remains neutral during an initial period of visual preprocessing time τ_{pre} . Then evidence is accumulated with an instantaneous evidence accumulation rate of mean δ (the drift rate) and standard deviation ς (the diffusion coefficient) via a Wiener process. The subject acquires either $\alpha = 1$ evidence unit or 0 evidence units to make a correct or incorrect decision respectively. After enough evidence is reached for either decision, motor response time τ_{motor} explains the remainder of that trial's observed reaction time. The 85th and 15th percentiles of Subject 10's single-trial drift rates $\delta_{i,10,1}$ in the low noise condition are shown as orange and green vectors, such that it would take 253 and 299 ms respectively to accumulate the .5 evidence accumulation units need to make a correct decision if there was no variance in the accumulation process. The larger drift rate is a linear function of the larger single-trial N200 amplitude (**), while the smaller drift rate is a linear function of the smaller N200 amplitude (*). The scalp activation (SVD weights multiplied by one trial's N200 amplitude) of this subject's response to the visual signal ranges from -13 to $13 \mu V$ on both trials. The two dark blue and red evidence time courses were randomly generated trials with the larger drift rate. The three dotted, light blue evidence time courses were randomly generated trials with the smaller drift rate. True Wiener processes with drifts $\delta_{i,10,1}$ and diffusion coefficient $\varsigma_{10,1}$ were estimated using a simple numerical technique discussed in Brown et al. (2006).

Figure 2: The time course of one trial of the experimental stimulus. One trial consisted of the following: 1) 750 ms of fixation on a black cross on a gray screen, then 2) visual contrast noise changing at 8 Hz for 750 ms while maintaining fixation (dubbed the *cue interval*) and 3) a circular field of small oriented bars flickering at 15 Hz overlaid on the changing visual noise for 1000 to 2000 ms while maintaining fixation (dubbed the *response interval*). The subjects' task was to indicate during the response interval whether the bars were on average oriented towards the "top-right" (45° from horizontal; as in this example) or the "top-left" (135° from horizontal). It was assumed that subjects' decision making process occurred only during the response interval but could be influenced by both onset attention to the visual noise during the cue interval and onset attention to visual signal during the response interval.

Figure 3: A visual representation of the singular value decomposition (SVD) method for finding single-trial estimates of evoked responses in EEG. The EEG presented here is time-locked to the signal onset during the response interval, such that the single-trial ERP encoded the response to the signal onset. A single trial of EEG from Subject 16 (Left) can be thought of as a time by channel ($T \times C$) matrix. The first SVD component explained the most variability (79.9%) in Subject 16's ERP response to the signal across all trials in the training set. SVD weights $\mathbf{v}$ ($C \times 1$) are obtained from the ERP response (i.e. trial-averaged EEG; $T \times C$) and can be plotted on a cartoon representation of the human scalp with intermediate interpolated values (Middle). This specific trial's ERP (Right) was obtained by multiplying the time series data from each channel on this trial by the associated weight in vector $\mathbf{v}$ and then summing across all weighted channels.

(a) Model 1

(b) Model 2

(c) Model 3

Figure 4: Graphical representations of the three hierarchical Bayesian models following the convention of Lee and Wagenmakers (2014). Each node represents a variable in the model with arrows indicating what variables are influenced by other variables. The magenta 2×1 vector of reaction time and accuracy $\mathbf{y}_{ijk}$ and the blue $(p + 1) \times 1$ vector of p EEG regressors (+1 intercept) $\mathbf{x}_{ijk}$ are observed variables, as indicated by the shaded nodes. Bolded blue variables indicate $(p + 1) \times 1$ vectors, such as the subject j level effects γ_{jk}^* of each EEG regressor and the condition k level effects $\mu_{(\gamma^*)_k}$ of each EEG regressor. In Model 3 for each trial i , values of non-decision time τ_{ijk} , drift rate (evidence accumulation rate) δ_{ijk} , the diffusion coefficient (evidence accumulation variance) ς_{ijk} are deterministic linear combinations of single-trial EEG regressors $\mathbf{x}_{ijk}$ and the effects of those regressors γ_{jk}^* that vary by subject and condition.

Figure 5: Single-trial evoked responses of an example subject, Subject 12, to the visual noise during the cue interval (Left) and single-trial evoked responses to the visual signal during the response interval (Right). Single-trial P200 and N200 magnitudes were found by finding peak amplitudes in 150 to 275 ms time windows (as indicated by the vertical dashed lines) of the SVD-biased EEG data in both the cue and response intervals. The first 300 ms of the intervals are sorted by single-trial P200 magnitudes in the cue interval and single-trial N200 magnitudes in the response interval. Latencies of the single-trial P200 and N200 components correspond to known latencies of P2 and N1 ERP components.

Figure 6: The posterior distributions of the effect of a trial's N200 latency during the response interval (onset attention latency to the signal stimulus) on trial-specific non-decision times τ_{ijk} for each subject j in the low noise condition $k = 1$. Subjects 2, 6, 7 and 11 were left out of the training set and their predicted posterior distributions are shown in red. Thick lines forming the distribution functions represent 95% credible intervals while thin lines represent 99% credible intervals. Crosses and horizontal lines represent posterior means and modes respectively. Also shown are the topographic representations of the channel weights of the first SVD component of each subject's response interval ERP, indicating the location of single-trial N200s over occipital and parietal electrodes. Evidence suggests that longer attentional latencies to the signal, N200 latencies, are linearly correlated with longer non-decision times in the low noise condition.

Figure 7: The posterior distributions of the effect of a trial's P200 amplitude during the cue interval (onset of attention to the noise stimulus) on trial-specific evidence accumulation rates δ_{ijk} for each subject j in the medium noise condition $k = 2$. Subjects 2, 6, 7 and 11 were left out of the training set, their predicted posterior distributions are shown in red. Thick lines forming the distribution functions represent 95% credible intervals while thin lines represent 99% credible intervals. Crosses and horizontal lines represent posterior means and modes respectively. Also shown are the topographic representation of the channel weights of the first SVD component of each subject's cue interval ERP, indicating the location of single-trial P200s over occipital and parietal electrodes. Evidence suggests that the effect of the attention to the noise, reflected in P200 amplitudes, positively influenced the drift rate of each subject in each trial, in the medium and high noise conditions.

Figure 8: Right and left sagittal and posterior views of localized single-trial P200 evoked potentials during the cue interval (Top) and localized single-trial N200 evoked potentials during the response interval (Bottom) averaged across trials and subjects. The cortical maps were obtained by projecting MNI-scalp spline-Laplacians (Nunez and Pilgreen, 1991; Deng et al., 2012) onto a subject's anatomical fMRI image via Tikhonov (L2) regularization, maintaining similar distributions of activity of the surface Laplacians on the cortical surface. Blue and orange regions in microamperes per mm^2 correspond to cortical areas estimated to produce negative and positive potentials observed on the scalp respectively. These two particular projections of the Laplacians suggest that P200 and N200 activity occurs in extrastriate cortices and areas in the parietal lobe.

Prediction of new data from <i>new</i> subjects				
		Model 1 Comparison	Model 2 EEG- δ, τ	Model 3 EEG- δ, τ, ς
Low	25 th $\mathbf{t}_1$ Percentile	−9.8%	22.1%	17.9%
	$\mathbf{t}_1$ Median	−37.6%	−11.6%	−14.3%
	75 th $\mathbf{t}_1$ Percentile	−67.3%	−46.7%	−47.6%
	Accuracy	−8.1%	−22.8%	−65.9%
Medium	25 th $\mathbf{t}_1$ Percentile	−5.3%	8.4%	5.3%
	$\mathbf{t}_1$ Median	−30.7%	−22.5%	−25.7%
	75 th $\mathbf{t}_1$ Percentile	−62.6%	−53.9%	−56.2%
	Accuracy	−13.1%	−35.5%	−67.7%
High	25 th $\mathbf{t}_1$ Percentile	−1.7%	8.2%	7.7%
	$\mathbf{t}_1$ Median	−12.0%	−4.3%	−2.9%
	75 th $\mathbf{t}_1$ Percentile	−37.6%	−25.5%	−23.7%
	Accuracy	−4.3%	−14.1%	−29.1%

Table 2: Percentage of across-subject variance explained by out-of-sample prediction (R^2_{pred}) for accuracy and summary statistics of new subjects’ and correct-RT distributions. Posterior predictive distributions were generated for 4 new subjects by drawing from condition level posterior distributions. Most R^2_{pred} measures are negative because the amount of variance in prediction was greater than the variance of the measure across subjects; however the relative values from one model to the next are still informative about the improvement in prediction ability. The model without single-trial EEG measures, Model 1, does not predict new subjects’ correct-RT distributions. Models *with* single-trial EEG measures of onset attention, Model 2 and Model 3, can predict some variance of the new subjects’ 25th percentiles, with Model 2 outperforming Model 3.

Prediction of <i>training</i> data from <i>known</i> subjects				
		Model 1 Comparison	Model 2 EEG- δ, τ	Model 3 EEG- δ, τ, ς
Low	25 th $\mathbf{t}_1$ Percentile	96.5%	97.1%	97.8%
	$\mathbf{t}_1$ Median	96.6%	96.5%	97.3%
	75 th $\mathbf{t}_1$ Percentile	91.4%	93.4%	94.5%
	Accuracy	95.1%	95.2%	97.3%
	$\mathbf{t}_0$ Median	-118.8%	-108.3%	-111.6%
Medium	25 th $\mathbf{t}_1$ Percentile	86.0%	87.5%	88.6%
	$\mathbf{t}_1$ Median	95.9%	95.6%	96.3%
	75 th $\mathbf{t}_1$ Percentile	84.7%	89.4%	90.1%
	Accuracy	90.7%	94.1%	95.3%
	$\mathbf{t}_0$ Median	-163.9%	-158.6%	-163.9%
High	25 th $\mathbf{t}_1$ Percentile	85.4%	87.3%	86.7%
	$\mathbf{t}_1$ Median	93.1%	92.5%	92.9%
	75 th $\mathbf{t}_1$ Percentile	79.1%	83.8%	84.0%
	Accuracy	95.9%	97.4%	95.9%
	$\mathbf{t}_0$ Median	-73.4%	-71.2%	-76.4%

Table 3: Percentage of variance across subjects explained by in-sample prediction (R^2_{pred}) for summary statistics of known subjects’ accuracy-RT distributions. All three models fit accuracy and correct-RT $\mathbf{t}_1$ data very well, explaining over 92% of median correct-RT and over 90% of accuracy in each condition. However none of the models explain incorrect-RT $\mathbf{t}_0$ distributions well, a known problem for simple diffusion models that can be overcome by including variable drift rates directly in the likelihood function (Ratcliff, 1978; Ratcliff and McKoon, 2008).