

A tutorial on fitting joint models of M/EEG and behavior to understand cognition

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

^a*Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*

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

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

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

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

Abstract

We present motivation and practical steps necessary to find parameter estimates of joint models of behavior and neural electrophysiological data. This tutorial is written for researchers wishing to build joint models of human behavior and scalp and intracranial electroencephalographic (EEG) or magnetoencephalographic (MEG) data, and more specifically those researchers who seek to understand human cognition. Although these techniques could easily be applied to animal models. Joint modeling of M/EEG and behavior requires some knowledge of existing computational and cognitive theories, M/EEG artifact correction, M/EEG analysis techniques, cognitive modeling, and programming for statistical modeling implementation. This paper seeks to give an introduction to these techniques as they apply to estimating parameters from neurocognitive models of M/EEG and human behavior, and to evaluate model results and compare models. Due to our research and knowledge on the subject matter, our examples in this paper will focus on testing specific hypotheses in human decision-making theory. However most of the motivation and discussion of this paper applies across many modeling procedures and applications.

Keywords: Computational modeling, Cognitive modeling, Electroencephalography (EEG), Magnetoencephalography (MEG), Psychology, Neuroscience

¹ 1. Motivation to model

² Fitting joint models of M/EEG and human behavior have been used to answer diverse
³ questions about cognitive topics such as working memory (Zhang et al., 2018), reinforcement
⁴ learning (Frank et al., 2015; Swart et al., 2018), cognitive abilities (Schubert et al., 2019),
⁵ and even the study of dyslexia in children (Manning et al., 2021). There is considerable
⁶ value to joint modeling in M/EEG measures to behavioural data to answer questions about
⁷ cognition. For instance, Schubert et al. (2019) show that neural processing speed, as reflected
⁸ in stimulus-locked EEG measures, describes variation in cognitive task performance across

*Corresponding author

Email address: m.d.nunez@uva.nl (Michael D. Nunez)

9 individuals. In another study, Nunez et al. (2015) show that EEG measures of attention can
10 reveal how individual differences in visual attention affect differentiable cognitive components
11 of decision-making.

12 What do we hope to achieve by finding parameter estimates of joint models of M/EEG
13 and human behavior? We could use these parameter estimates to draw conclusions about
14 a scientific hypothesis, to help differentiate between theories, to find models that best pre-
15 dict human behavior and brain dynamics, or to teach students how to fit models to data.
16 While the possible goals that could be realized by joint modeling of M/EEG and human
17 behavior are numerous (e.g. see Kording et al., 2018), most researchers who fit joint mod-
18 els will seek to either (1) test specific hypotheses or differentiate theories in fields such as
19 Neuroscience and Psychology or (2) maximize prediction of M/EEG signals and/or human
20 behavior. Sometimes these two goals can be simultaneously realized, but maximizing pre-
21 diction of M/EEG and human behavior is often best achieved with atheoretical approaches
22 based in Machine Learning (ML) or Artificial Intelligence (AI). ML and AI are best used in
23 scenarios where maximizing prediction is most important and understanding the cognitive
24 process is not critical, for instance in many Brain Computer Interfaces (BCIs).

25 In this tutorial we will focus primarily on joint modeling for testing specific hypotheses.
26 The diversity in *methods* used to perform joint modeling of M/EEG and behaviour is large.
27 The studies mentioned above include a variety of methods to perform joint modeling. There
28 are also many related studies that correlate cognitive model parameters to observed M/EEG
29 measures or correlate observed behavioural measures to computational parameters of EEG
30 (e.g. O'Connell et al., 2012; Gluth et al., 2013; Jagannathan et al., 2021). Our intention
31 in this tutorial not to review all cognitive topics (e.g. see Hawkins et al., 2022) nor joint
32 modeling techniques of M/EEG and behaviour (e.g. see Palestro et al., 2018). Instead we
33 will focus on clarifying common modeling examples, EEG data collection and analysis, and
34 tools to implement joint models. We will cover experimental design, M/EEG analysis and
35 behavioral analysis techniques necessary for joint modeling, as well as the modeling itself.
36 Due to our research and knowledge on the subject matter, our examples in this paper will
37 focus on testing specific hypotheses in human decision-making theory. However all techniques
38 and software presented here can be applied to testing any formal hypotheses involving the
39 relationship of M/EEG to human cognition and behavior.

40 *1.1. One example topic: Decision-making*

41 One topic of interest is whether specific M/EEG signals reflect cognitive components of
42 decision-making. This has been the focus of a large body of previous work (e.g. see O'Connell
43 et al., 2018) including our own work (see **Figure 1**, Lui et al., 2021). This work has lead
44 to specific testable questions that can be answered with joint modeling of M/EEG data,
45 response time data, and choice data from tasks in which participants make simple decisions.
46 However before we discuss this example joint modeling work, let us first briefly review a key
47 theory of decision making.

48 Sequential sampling models assume that humans and animals accumulate evidence for a
49 particular choice over the course of a decision by *sampling* from external or internal evidence.
50 This *evidence* for a decision is usually considered a cognitive representation or a direct neural
51 representation (e.g. changing firing rates of neurons over time). Simulating and fitting
52 sequential sampling models are particularly useful for understanding quick decisions on the

53 scale of seconds. These models make predictions about the time course of decision-making,
54 while other decision-making models, such as Signal Detection Theory (SDT), do not make
55 any predictions about the time course of decision-making. Drift-diffusion models (DDMs) are
56 a particular class of sequential sampling models that assume a Wiener process of evidence
57 accumulation. A Wiener process of evidence accumulation is random walk process with
58 an infinitesimal (infinitely small) time step (see middle of **Figure 1**). Mathematically, the
59 evidence E_t at time t follows a Wiener process with drift rate δ and accumulation variance
60 ς^2 (Ross, 2019) such that

$$E_t \sim \text{Normal}(\delta t, \varsigma^2 t) \quad (1)$$

61 Note though that DDMs and associated model-variants are often used due to the models'
62 mathematical utility, rather than a theoretical belief that a time step of evidence accumula-
63 tion should be infinitesimal in the brain.

64 Sequential sampling theory leads naturally to specific testable questions that can be
65 answered with joint modeling of M/EEG data. Do time-averaged event-related potentials
66 (ERPs) encode the demarcation point between visual encoding and evidence accumula-
67 tion (Nunez et al., 2019a)? Do motor preparation signals over the motor cortex track evidence
68 accumulation time (Lui et al., 2021)? How do EEG measures of visual attention affect
69 the decision-making process, and in what precise way does visual attention affect different
70 computational components of decision-making (Nunez et al., 2015, 2017)? There exist mul-
71 tiple analysis methods to help answer these questions with EEG and behavior (see Bridwell
72 et al., 2018). However we have preferred to use the implementation of joint modeling of
73 EEG and human behavior to understand data from participants who performed hypotheses-
74 differentiating experiments.

75 *1.2. One example question: Do EEG signals encode a sequential-sampling of evidence?*

76 One more specific question is whether EEG signals encode a sequential-sampling of ev-
77 idence, such has been found in single neurons and neural populations within intracranial
78 recordings of the lateral intraparietal (LIP) cortex and superior colliculus (SC) of the macaque
79 brain during single experimental trials (Roitman and Shadlen, 2002; Shadlen and Kiani, 2013;
80 O'Connell et al., 2018; Jun et al., 2021). EEG signals time-locked to specific events such as
81 the onset of a visual stimulus, e.g. the P300 / Centro-Parietal Positivity (CPP) waveform,
82 and EEG signals time-locked to the response, e.g. the Readiness Potential (RP), have been
83 proposed to be related to evidence accumulation and the timing of decisions (O'Connell
84 et al., 2012; Gluth et al., 2013; Twomey et al., 2015; van Ravenzwaaij et al., 2017; Lui et al.,
85 2021). We expect future joint modeling work will further help differentiate whether these
86 signals are exactly encoding evidence accumulation, correlated processes, or mixtures of sig-
87 nals (Philiastides et al., 2014). Related questions that can be answered with joint modeling
88 work are: (1) what EEG preprocessing steps and recording procedures should be used to
89 best extract evidence accumulated related signals? and (2) for what specific conditions and
90 task paradigms these signals encode evidence accumulation? A table of related questions
91 that can be answered with neurocognitive models using DDMs can be found in **Table 1**.
92 In this paper we show how combining these specific EEG signals and behavioural data in
93 neurocognitive modeling will lead to better and more extensive knowledge of individual dif-
94 ferences and single-trial estimates of human cognition. In our example methods presented in

95 this paper we build models and present model fitting procedures that can best answer the
96 question of whether EEG signals encode a sequential-sampling of evidence.

97 **2. Models to describe joint data**

98 *2.1. The basic terms of modeling*

99 *Cognitive models* include parameters of psychological processes that describe human (or
100 animal) behavior. These models are often used to describe human and animal behavior in
101 psychological experiments or natural environments, and these models are often developed by
102 researchers in the scientific field of Mathematical Psychology. A *parameter* of a cognitive
103 model is a variable that can take a pre-specified range of values that describe data. Although
104 note that multiple parameters of a model are usually required to describe data. The param-
105 eters of cognitive models often directly relate to unobserved psychological concepts such as
106 memory capacity or general cognitive ability (Lee et al., 2019; Schubert et al., 2019). Sig-
107 nal Detection Theory (SDT) that explains choice and accuracy data could be considered a
108 cognitive model since it contains two parameters that describe both the ability and choice
109 bias of a human participant (Hautus et al., 2021). Another example are Drift-Diffusion Mod-
110 els (DDMs) of choices and response time (RT) data during human decision-making, which
111 contain cognitive parameters that describe speed-accuracy trade-offs, speed of evidence accu-
112 mulation for one choice or another, and decision biases (Ratcliff et al., 2016). The cognitive
113 interpretations of parameters of new cognitive models should be tested by experimentation
114 with differentiating experimental conditions (discussed below). However some widely-used
115 cognitive models, such as SDT and DDMs, have parameters whose cognitive interpretations
116 are now widely accepted by researchers due to the results of multiple experimental stud-
117 ies. For instance, Voss et al. (2004); Dutilh et al. (2019) generally found that parameters of
118 DDMs (namely speed-accuracy tradeoff parameters, non-decision time parameters, evidence
119 accumulation rate parameters, and evidence bias parameters) that describe human choice
120 and reaction time are all manipulated in expected directions by proper experimental con-
121 ditions. Although these experimental manipulations are not perfect (Dutilh et al., 2019).
122 These prior results allow us to draw new conclusions in new data while assuming some of
123 these parameters map onto the expected cognitive function.

124 Cognitive models are sometimes defined separately to *computational models*. Computa-
125 tional models often focus on modeling of brain dynamics with parameters that have specific
126 neural correlates (Blohm et al., 2020; Glomb et al., 2021). Typically computational models
127 are used in the scientific field of Computational Neuroscience. Computational models do not
128 necessarily have parameters with interpretable psychological processes. Although because
129 computational models sometimes describe human behavioral data, the term “computational
130 model” is sometimes used to also include cognitive models (e.g. Wilson and Collins, 2019).

131 *Neurocognitive models* are joint models of brain activity and human behavior that seeks to
132 combine cognitive modeling with necessary links between (1) brain dynamics as measured by
133 or derived from M/EEG, (2) cognition and other psychological concepts expressed as formal
134 models, and (3) human behavioral data. As a generalization, with joint modeling we seek
135 to understand how macro-level neurophysiology (as measured by scalp-recorded EEG, MEG,
136 or even depth-recorded EEG) encodes human cognition which gives rise to human behavior.
137 For instance, evidence accumulation is a cognitive process during decision-making, but may

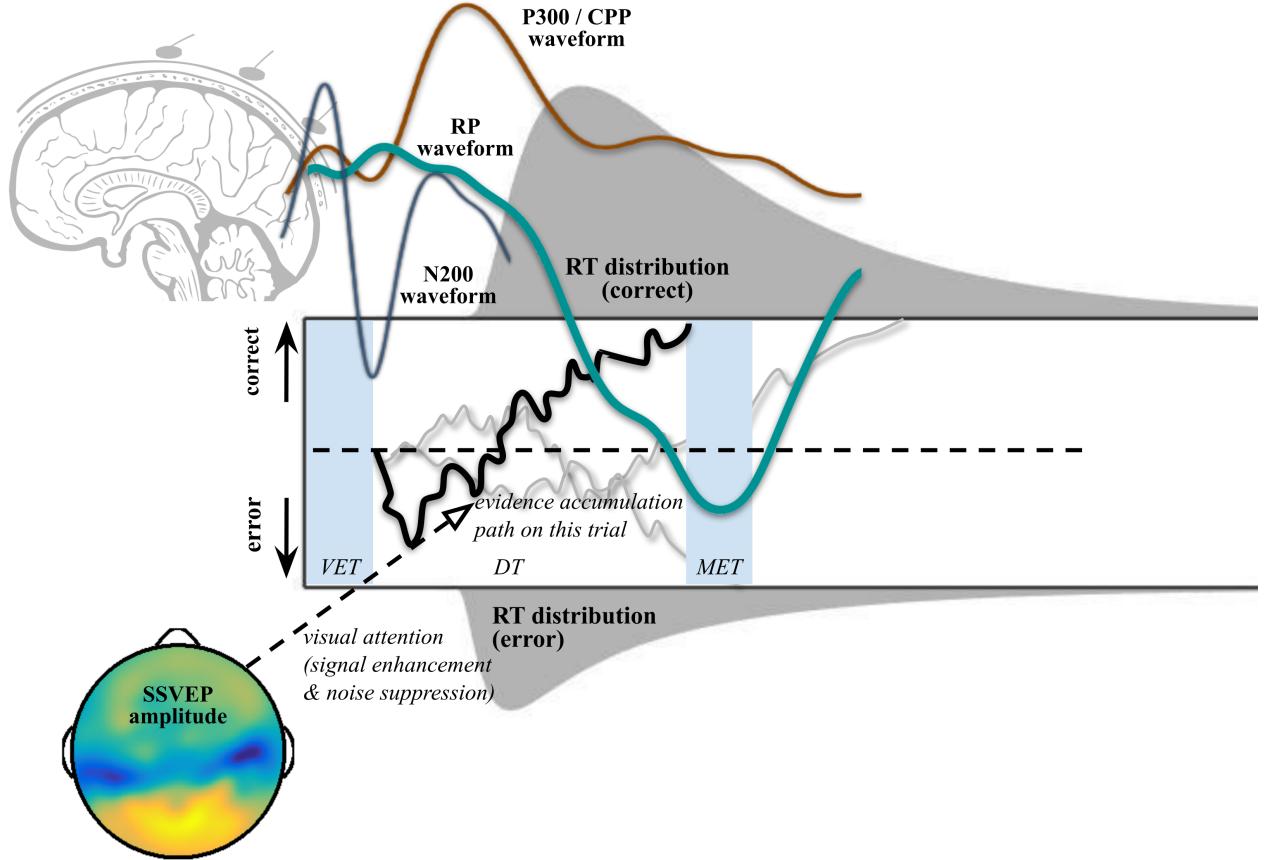


Figure 1: A theoretical representation of some modeling studies to discover cognitive mechanisms of decision-making using *neurocognitive* modeling of EEG and human behavior during decision-making tasks. Bold text represents observed data (EEG measures or human behavioral data) while italic text represents derived cognitive parameters that can be estimated through joint modeling. Event-Related Potentials (ERPs; represented by 3 waveforms beginning at the cartoon image of the brain in the top right) and frequency-domain EEG measures (bottom left: EEG amplitudes that were spline-interpolated between electrodes on a flat representation of the human scalp) have been used in joint modeling to understand human cognition in the context of Neural Drift-Diffusion Models (NDDMs). Human behavioural data such as choice-RTs (response time distributions shown for correct responses, top, and error responses, bottom flipped) are also used to fit NDDMs and Drift-Diffusion Models (DDMs). In NDDMs, like DDMs, correct and error responses are described after enough cognitive evidence is reached, represented by the cognitive evidence accumulation passing one of two boundaries during decision time (this trial represented as a black line with two other grey lines representing other simulations from the same process that describe response times and possibly EEG potentials). Particular ERPs of interest are N200, P300 / CPP, and RP waveforms. N200 waveforms are thought to reflect visual encoding time (VET) and the onset of evidence accumulation (Nunez et al., 2019a). The P300 or Centro-parietal positivity (CPP) are thought to reflect decision time (DT) and possibly the evidence accumulation process itself (O'Connell et al., 2012; Kelly and O'Connell, 2013; O'Connell et al., 2018; van Ravenzwaaij et al., 2017). The Readiness Potential (RP) is a motor related preparatory signal thought to reflect DT and motor execution time (MET) under certain experimental conditions (Lui et al., 2021). Steady-state visual evoked potentials (SSVEPs) can be used to estimate visual attention and in particular, signal enhancement and noise suppression that could affect the rate and variance of evidence accumulation (Nunez et al., 2015). A table of related questions and other neurocognitive work using DDMs can be found in **Table 1**.

138 also have direct neural correlates in EEG (Forstmann et al., 2016; O’Connell et al., 2018;
139 Lui et al., 2021), and thus drift rates of DDMs could describe choice-RT distribution shapes,
140 evoked EEG potentials, or both simultaneously. Thus these *neurocognitive* models can be
141 used to develop and test theories in both Psychology and Neuroscience. Specifically we focus
142 in this paper on how modeling can be used to directly test hypotheses which involve observed
143 M/EEG dynamics, human cognition, and human behavior.

144 When *simulating* models, parameters are fixed to certain values and a model generates
145 fake data that could be compared to real data. For instance, a neural drift diffusion model
146 (NDDM) with multiple user-defined parameters could generate fake (but informative) EEG
147 potentials, choices, and response times. To *fit* a model is to discover a set of parameter
148 estimates, or parameter uncertainties using Bayesian methods, that best describe known
149 data given the model architecture and assumptions. Fitting models to data is a useful
150 method to test hypotheses by either (1) by directly estimating and then evaluating parameters
151 (e.g. compare parameter estimates across experimental conditions) or (2) comparing multiple
152 models’ ability to describe data.

153 *2.2. Translating neurocognitive theory into mathematical models*

154 Often modeling involves simplifying the broader mathematically-defined theories of the
155 brain and human behavior in order to fit data accurately and efficiently. In this way theory,
156 or an approximation of theory, can be used to answer questions directly. Fitting models
157 to EEG and behavioral data requires (1) knowledge of cognitive or neurocognitive theory,
158 and then (2) work to quantitatively formalize the hypotheses to be tested in the context
159 of theory. Good theory should be quantitatively defined, and the best theory is precisely
160 mathematically defined (Guest and Martin, 2021). Although many other qualifications may
161 be needed for good theory, which are discussed elsewhere (e.g. van Rooij and Baggio, 2020).

162 A model should be able to be written as a series of mathematical equations that describe
163 the data. Let us assume that a participant, Roos, wore an EEG cap while playing a simple
164 video game where she made a correct or incorrect answer on multiple trials of the game.
165 Researchers extracted EEG and behavioral data from the data collection hardware and are
166 now interested in how the Centro-Parietal Positivity (CPP) rise over time within a trial (e.g.
167 the CPP “slope”) influences accuracy and response times (RTs). Note that the CPP slope
168 is a signal in the EEG that is found after the onset of visual stimuli, and is thought to
169 be a reflection of the computation of visual evidence in the brain (O’Connell et al., 2012;
170 O’Connell et al., 2018). Specifically the researchers are interested in how three measures
171 from Roos are related: CPP slopes, accuracy, and RTs. The researchers obtain observations
172 of each of the three measures on every trial of the game. One simple joint model could just
173 assume that RTs and accuracy are influenced by the CPP slope. Thus a simple joint model
174 (**Model 1**) would just be two equations, linear regression and logistic regression. Let CPP
175 slope be denoted by the variable c , accuracy be denoted by x , and response times be denoted
176 by r . Note that accuracy x can be either 0 or 1, and that x , r , and c can vary on every trial i
177 and by participant j . The parameters of the model can change by participant j . Researchers
178 can then fit this model to Roos’ data as well as other participants.

$$r_{ij} \sim Normal(\theta_{0j} + \theta_{1j}c_{ij}, \sigma_j^2) \quad (2)$$

$$x_{ij} \sim Bernoulli(p_j) \quad (3)$$

$$\log\left(\frac{p_j}{1-p_j}\right) = \gamma_{0j} + \gamma_{1j}c_{ij} \quad (4)$$

179 Note that we picked the logit (“log-odds”) function $logit(p) = \log(p_j/(1-p_j))$ within the
 180 logistic regression framework, although we could use any function that maps probabilities
 181 bounded from 0 to 1 to the continuous $(-\infty, \infty)$ scale. The θ and γ parameters in this model
 182 provide the relationship of EEG measure to behavior. Note that we can test whether the non-
 183 intercept θ_1 and γ_1 parameters are near zero in order to test many hypothesized relationships
 184 between CPP slope and behavior. Alternatively, we could compare the fit of the behavioral
 185 data in this model to a model where the EEG does not influence the data generators. In a
 186 comparison model, response times and accuracies would be described only by the mean and
 187 probability parameters (e.g. setting θ_1 and γ_1 both to 0 before model fitting). Note that
 188 **Model 1** does not assume EEG reflects any *particular* type of cognition or computation.
 189 However our example question is more specific about the type of computation and cognition
 190 the CPP slope could reflect. Namely, that CPP slope reflects evidence accumulation speed.
 191 **Model 1** assumes that any relationship is linear between CPP slopes and response times as
 192 well as CPP slopes and log odds of accuracies. Fitting **Model 1** will likely yield information
 193 about the existence of a CPP-behavior relationship because linear and logistic regression are
 194 useful in finding relationships where there is *some* true relationship between variables. **Model**
 195 **1** would be particularly helpful if we wanted to test if there was any possible relationship in
 196 an exploratory analysis.

197 Researchers instead might want to directly test whether the CPP slopes reflect specific
 198 cognitive components of decision-making that describe both the accuracy and response time
 199 jointly. A second model (**Model 2**) could then test simple linear relationships to cognitive
 200 parameters. In particular we might be interested in how CPP slopes describe the drift-rate
 201 parameters δ and the non-decision time parameters τ . The drift rate δ reflects the mean
 202 rate of evidence accumulation for each infinitesimal time step of a Wiener process (Ratcliff
 203 et al., 2016). Non-decision time τ on every trial is any time in a response time not due to
 204 a Wiener process (such as visual encoding time; VET). Lets assume that δ and τ can vary
 205 on every experimental trial. We also assume that other parameters of a DDM do not vary
 206 with CPP slope and are fixed across trials, namely: the boundary separation parameter α_j
 207 which describes the amount of relative evidence to make a correct choice, the initial bias
 208 parameter β_j towards the correct choice before evidence accumulation occurs, and trial-to-
 209 trial variability in drift rate, η_j , that is not due to CPP slope variability. This DDM describes
 210 both response times r and choices x per participant j and trial i .

$$(r_{ij}, x_{ij}) \sim DDM(\delta_{ij}, \tau_{ij}, \alpha_j, \beta_j, \eta_j) \quad (5)$$

$$\delta_{ij} = \xi_{0j} + \xi_{1j}c_{ij} \quad (6)$$

$$\tau_{ij} = \lambda_{0j} + \lambda_{1j}c_{ij} \quad (7)$$

Note that the parameters ξ and λ have different meanings in this model with embedded cognitive components compared to the θ and γ parameters in the previous model. The parameters ξ and λ are the intercept and effect parameters of the CPP slopes on the evidence accumulation rate and non-decision time respectively. We will call this model a *neurocognitive* model because it assumes particular types of cognition during decision making and contains neural data, in addition to behavioral data. We have fit this class of model, which assumes single-trial EEG measures describe single-trial DDM parameters, in previous work (Nunez et al., 2017, 2019a).

Another class of neurocognitive model is one in which the EEG measures themselves are described by the cognitive parameters. For instance, we can directly test the underlying computational role of the CPP slope in cognition by testing how well the following model fits the data once parameters are estimated and how well this model predicts new data. In **Model 3**, the mean of each trial's CPP slope c is described by the drift rate for each participant j .

$$(r_{ij}, x_{ij}) \sim DDM(\delta_j, \tau_j, \alpha_j, \beta_j, \eta_j) \quad (8)$$

$$c_{ij} \sim Normal(\delta_j, \sigma_j^2) \quad (9)$$

such that five cognitive parameters of a DDM, δ , τ , α , β , and η vary by participant j . Note that one of those cognitive parameters, the mean rate of evidence accumulation across trials δ is also a computational parameter that describes the CPP slope on every trial i and each participant j . There is one more additional computational parameter in this model that describes the observation noise of the CPP slope on every trial i , parameter σ for each participant j . This model is similar to, although more specific than, a model previously used by van Ravenzwaaij et al. (2017).

Better neurocognitive models might extend **Model 3** by describing the CPP slopes from a function of multiple cognitive parameters, for instance. Other models could contain computational parameters that reflect brain dynamics that are described by cognitive parameters. Different modeling strategies are discussed by Turner et al. (2017). All of the example models and other joint models can be simulated through the equations listed and by choosing a set of values for parameters. We can build a variety of neurocognitive models to test specific theories.

2.3. The importance of model simulation

Model simulation is usually the first step in generating new models, and can be helpful in understanding the neurocognitive theory. Model fitting intrinsically makes many assumptions about the data. Joint modeling of M/EEG and behavior will make assumptions about cognitive and computational processes in the brain and implicit assumptions about what cognitive and computational processes are **not** occurring. These assumptions will affect parameter fitting results significantly. It is therefore imperative that the researcher understand what assumptions they are making and how those assumptions can be violated.

The best way to explore these assumptions is through the two-step process of (1) simulation of models using different model types and realistic parameter ranges and then (2) fitting all the simulated data using the specific model and fitting procedure to be used in the

Table 1: A table of example hypotheses that could be tested directly using neuro-cognitive Drift Diffusion Models (DDMs) of M/EEG and behaviour. Each hypothesis is derived directly from existing literature in the fields of Cognitive Neuroscience and model-based Cognitive Neuroscience. Each *cognitive mechanism* is associated with a specific *computational role* in the human brain that is measured with a M/EEG *neural signature* that is hypothesized to be reflected in the relationship with a cognitive parameter of a DDM. δ refers to the drift rate parameter, τ refers to the non-decision time, τ^v refers to visual encoding time (a component of non-decision time) τ^m refers to motor execution time (another component of non-decision time) α refers to the boundary separation parameter, β refers to the initial bias parameter, ς refers to the diffusion coefficient (undiscussed in this text but discussed in Ratcliff et al., 2016; Nunez et al., 2017, and elsewhere).

Cognitive Mechanism	Computational Role	Neural Signature	DDM Parameters	References
Visual evidence accumulation	Evidence accumulation rate	P300 slopes	δ	(Philiastides et al., 2006; Ratcliff et al., 2009; Philiastides et al., 2014; Twomey et al., 2015; van Ravenzwaaij et al., 2017; Kohl et al., 2020)
Figure-ground segregation	Visual encoding time (VET)	N200 latencies	τ / τ^v	(Loughnane et al., 2016; Nunez et al., 2017, 2019a; Ghaderi-Kangavari et al., 2021, 2022)
Motor execution	Motor execution time	Beta (15-25 Hz) desynchronization	τ / τ^m	(Crone et al., 1998; McFarland et al., 2000)
Motor cortex preparation	Motor evidence accumulation	Readiness Potentials	δ	(Gluth et al., 2013; Lui et al., 2021)
Speed-accuracy tradeoff	Changing neural threshold	Theta (4-7 Hz) power	α	(Cavanagh et al., 2011; Frank et al., 2015)
Strategy adjustment	Changing neural threshold	Contingent Negative Variation	α	(Boehm et al., 2014)
Prestimulus activation	Bias	Occipital EEG amplitude to predict choice	β	(Bode et al., 2012)
Visual attention	Variability in evidence (diffusion)	Steady-state visual evoked potentials	δ, ς	(Nunez et al., 2015; Rangelov and Mattingley, 2020)
Attentional gating	Internal neural noise	Alpha (8-12 Hz) power	ς	(Pfurtscheller et al., 1996; Jensen and Mazaheri, 2010; Klatt et al., 2020)

analysis of real data. Simulations can show the researcher under which conditions the fitting procedure can break or produce spurious results. Simulation of models, as opposed to fitting of models, is also one way to formally define the underlying *theory* to be tested, in that it is quantifiable, and perhaps complex, but rigorously defined (Guest and Martin, 2021).

To simulate joint models of M/EEG, it is best to write the model using existing statistical libraries with coding languages such as Python or R. We reproduced one possible simulation of **Model 3** here. A code snippet of **Model 3** is provided in **Code snippet 1** with associated full Python code located at https://github.com/mdnunez/pyhddmjags/blob/master/simpleCPP_sim.py as of March 2022. We first generate random values of all parameters for one participant. We then simulate data for all trials i for this participant from those parameters within the two equations of **Model 3** in a for loop (shown in **Code snippet 1**). Our code simulates the model directly from an approximation of the Wiener process. This is especially useful for observing simulations of the evidence paths on every trials (see top left of **Figure 2**). Our code also simulates the CPP observations themselves from a sine wave with 1/4 of the period of the sine wave being the CPP slope (see bottom left of **Figure 2**).

Code snippet 1: Python code from simulation of **Model 3**

```
266 plot_time = np.linspace(0, step_length * nsteps, num=nsteps)
267 for n in range(0, ntrials):
268     random_walk = np.empty(nsteps)
269     drift = stats.norm.rvs(loc=delta, scale=eta)
270     cpp_slopes[n] = stats.norm.rvs(loc=delta, scale=sigma)
271     CPPs[:, n] = np.sin(2 * np.pi * ((cpp_slopes[n] / 4) * (plot_time - ndt)))
272     random_walk[0] = beta * alpha
273     for s in range(1, nsteps):
274         random_walk[s] = random_walk[s - 1] + stats.norm.rvs(
275             loc=drift * step_length, scale=varsigma * np.sqrt(step_length))
276         if random_walk[s] >= alpha:
277             random_walk[s:] = alpha
278             rts[n] = s * step_length + ndt
279             choice[n] = 1 # Correct choice shown with positive RTs
280             break
281         elif random_walk[s] <= 0:
282             random_walk[s:] = 0
283             rts[n] = s * step_length + ndt
284             choice[n] = -1 # Incorrect choice shown with negative RTs
285             break
286         elif s == (nsteps - 1):
287             rts[n] = np.nan
288             choice[n] = np.nan
289             break
```

When simulating a model it is useful to plot elements of the model itself. For instance in **Figure 2** we have plotted dynamics of the model itself, namely the CPP waveforms on single-trials and the evidence paths themselves. We have also plotted the distributions of both the choice response times and the CPP slopes. We can also choose to plot other diagnostics such as the cumulative distribution function, and return specific statistics about our simulated

295 data (e.g. mean, median, maximum, and minimums). In this way we can make sure our
296 model, and quantitatively defined theories, are logical and could describe real data. Note
297 however that we may not observe everything about our model in real data nor be able to
298 estimate everything about our model from the data. For instance, we may never observe the
299 cognitive evidence paths themselves (top right of **Figure 2**) in real data. Later, after we
300 discuss model fitting, we will also discuss how we can also use simulations to test parameter
301 recovery. We will show that *fitting Model 3* to data is indeed useful.

302 3. Experimental manipulations and experimental design

303 The goal of experimentation should be to design experimental conditions that best answer
304 scientific questions. In this section we will focus on experiments that are optimally designed
305 to produce data for use with joint modeling. One particular *neurocognitive* theory suggests
306 that a particular trial-averaged EEG signal, the CPP, in response to the onset of a picture
307 on a computer screen is expected to be a neural signature of evidence accumulation during
308 decision-making based on that picture. For example, the CPP reflects evidence accumulation
309 to decide whether a noisy picture is a face or a car (e.g. Ostwald et al., 2012). This is a theory
310 suggested by experimental and theoretical work by O’Connell et al. (2012); O’Connell et al.
311 (2018), previously mentioned. Thus we should design experiments to collect M/EEG and
312 behavioral data and/or pick an existing data set that (1) best test the hypothesis and (2)
313 test the limits of this hypothesis.

314 3.1. Hypothetical Experiment 1

315 We know that drift rates in perceptual decision making tasks are affected by perceptual
316 difficulty (Voss et al., 2004; Dutilh et al., 2019). Therefore if CPPs are signatures of the
317 computational mechanism of evidence accumulation in the brain, we expect drift rates and
318 CPP slopes to both change across perceptual difficulty conditions, and specifically maintain
319 the same relationship across many experimental conditions of various difficulty. For instance
320 participants could perform a Random Dot Motion (RDM) task with many different levels
321 of perceptual difficulty across trials, a replication of work by Kelly and O’Connell (2013).
322 A RDM task is a task in which a field of moving dots appears on each experimental trial
323 with a (typically small) percentage moving of dots moving in a specific direction (e.g. see
324 Newsome and Pare, 1988; Gherman and Philiastides, 2018). Typically participants must
325 choose between one of two directions. If all the dots moved together on the screen, the
326 task would be too easy and would not result in different performance across conditions, not
327 optimally testing our neurocognitive theory. Therefore the other percentage of dots typically
328 move completely at random with no orientation information. The percentage *coherence* of
329 dots describes the percentage of dots that move in the correct orientation. Therefore let us
330 imagine a task in which trials are intermixed with 32%, 16%, 8%, 4%, and 0% coherence
331 values. Note that the difficulty of the coherence values will depend, among other factors, on
332 the size of the dots in the stimulus and the size of the entire stimulus for the participant.

333 We could fit the data from Experiment 1 to **Model 2** for each experimental condition
334 k , such that now each parameter and data type also varies by condition k . Comparing
335 parameter estimates of parameter the slope parameter ξ_{1jk} across the two conditions can
336 then test whether there is a fixed linear relationship between the slope of the CPP and the

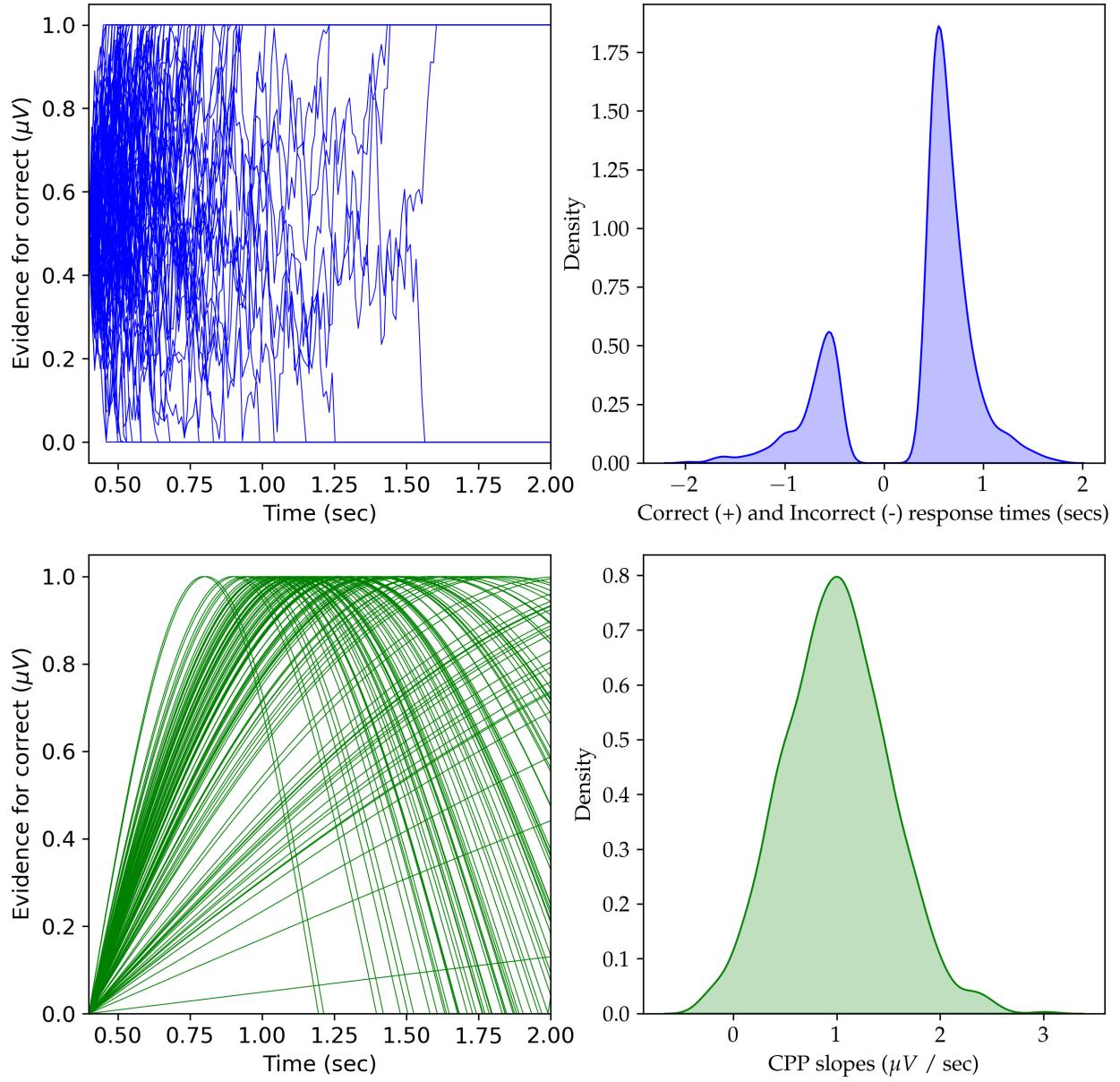


Figure 2: Diagnostic plots of the simulation of **Model 3**. We simulated **Model 3** using the provided Python code with 1000 trials for one participant. The top left figure is the simulated evidence accumulation paths from a Wiener process that reaches one of two boundaries to make a decision. The top right figure is the estimated density of incorrect and correct response times, with incorrect response times plotted as negative response times. The bottom left figure is the simulated CPPs on single-trials using a simple sine wave. The bottom right figure is approximated density of the simulated CPP slopes. This density should be approximately normal because of our modeling assumption for CPP slopes.

drift rate in all experimental conditions, even when other parameters such as the drift-rate itself δ changes across experimental conditions. Specifically we could develop statistical tests to test whether $\xi_{1j1} = \xi_{1j2} = \xi_{1j3} = \dots = \xi_{1jK}$ from the resulting parameter estimates in all participants, for instance, we could analyze posterior distributions of these parameters to calculate posterior probability or Bayes Factors (see below). Alternatively we could compare how two models explain the data and predict new data. For instance, we could compare **Model 2A** to **Model 2B**, where **Model 2A** has fixed effect parameters ξ_{1j} across conditions and **Model 2B** has effect parameters ξ_{1jk} that are free-to-vary across conditions.

3.2. Hypothetical Experiment 2

The first hypothetical experiment demonstrates a simple perceptual manipulation intended to drive cognition that best answers our hypothesis. In the second hypothetical experiment we propose an intervention indented to drive the brain response using transcranial direct current stimulation (tDCS). While Experiment 1 is a well established experimental manipulation, the intervention in Experiment 2 may or may not have any effect on the participants' brain response, cognition, and behaviour. However Experiment 2 is a useful example to show that proper control conditions are often necessary to test a hypothesis using joint modeling.

Because of the aforementioned theory, we might expect transcranial direct current stimulation (tDCS) during a visual decision-making task to affect both CPP slopes c and evidence accumulation rate parameters (i.e. *drift rates*) δ in DDMs estimated from human behavior. A strong hypothesis is that we expect tDCS to affect both CPP slopes and drift rates equally since the theory is that the CPP is a signature of the computational mechanism of evidence accumulation in the brain. Thus tDCS could test the limits of the theory of the CPP reflecting evidence accumulation. We should at least have both (1) an experimental condition and (2) a proper control condition in which stimulating tDCS electrodes on the participant's head.

In tDCS work an experimental control is often a sham condition in which tDCS is turned on then off after a ramping period before a block of experimental trials. This sham condition seeks to achieve the sensation of tDCS stimulation by the participant for a block of experimental trials, but to not actually stimulate during those trials (e.g. see Au et al., 2021). We propose an experimental design with one experimental condition (1) in which the experimenters stimulate brain areas expected to be involved in decision-making (perhaps placing stimulation electrodes on the scalp over parietal cortices), and a control condition (2) in which experimenters stimulate brain areas not expected to be involved in decision-making, say placing tDCS electrodes on the scalp over the temporal cortex. We could also include a different sham condition, another experimental control, (3) in which tDCS electrodes do not stimulate the brain but current is still injected into the body, for instance placing and activating tDCS electrodes over the neck musculature in the back of the head.

We could again fit **Model 2B** to this data and analyze the results in a similar way to the previous experiment. For instance, suppose we fit the data to **Model 2B** and subsequently observe that both CPP slopes c and drift rates δ are significantly increased in condition (1) compared to sham condition (3), and that the two effect parameters of the model are equal, $\xi_{1j1} = \xi_{1j3}$. We also observe that only drift rates δ are significantly increased in condition (2) compared to the sham condition (3) and CPP slopes c are the similar in both conditions,

resulting in $\xi_{1j1} = \xi_{1j3} > \xi_{1j2}$. These results would be evidence that the CPP slope reflects evidence accumulation only in specific conditions. For instance, this might suggest the the CPP only reflects visual evidence accumulation in the parietal cortex, and is not the brain-wide cognitive phenomena thought to be reflected in decision-making behavior. Of course, we would need further experimentation to test this new, more specific theory. We should also build a new joint model to better reflect our new neurocognitive theory.

3.3. Theoretically informed experiments

While we should be able to *simulate* joint models that account for many experimental designs, experimental design choices may also be made *in preparation* for *fitting* joint models. The data from some experimental designs are more easily modelled due to (1) larger bases of research knowledge for some specific theories of human cognition and brain signals and (2) the availability of algorithms and software packages that allow fitting certain classes of models and data. That is not to say that some experimental and modeling work should not occur, only that more theoretical development, technical expertise, and/or model-fitting algorithm development is needed to answer some questions. Note that the idea of choosing experiments based on current states of theoretical knowledge and algorithm availability may be somewhat distasteful to researchers who feel that theory and algorithm development should occur to explain any collected data. However each of these two intermediate steps likely require extensive research (e.g. see Guest and Martin, 2021, for a discussion on theory development).

Decision-making data resulting from two-alternative forced choice (2AFC) experimental tasks, or similar tasks, are known to be easily studied using signal detection theory and sequential sampling models. However only some work has been performed to develop models for more complex decisions such as describing choices and reaction times for more than two alternatives (van Ravenzwaaij et al., 2020) or in continuous space (Ratcliff, 2018). This means that choosing an experimental design where participants perform a 2AFC task instead of a multi-alternative forced choice (MAFC) task will result in more theoretical understanding and programs to fit the data to joint models in order to study neurocognitive theory. For instance, given the current state of knowledge, if a researcher only cares about a scientific question that does not depend on MAFC versus 2AFC tasks, than the researcher might choose the 2AFC task because of the current state of model fitting procedures. However in the future it is likely that MAFC model fitting procedures will be more developed and widely used. In the future we expect model fitting procedures will also be more flexible in their implementation (e.g. see Radev et al., 2020). Furthermore if the goal is to develop new methods and/or to test something specific about multiple alternatives, than a MAFC experimental task would be best.

Some M/EEG signals will be more easily found (e.g. differentiated from other signals) in certain experimental conditions due to prior knowledge from a wealth of literature about these signals. For this reason the same experimental tasks are used often in M/EEG research and electrophysiology across experiments, such as the Random Dot Motion (RDM) task (e.g. see Newsome and Pare, 1988; Gherman and Philiastides, 2018). For instance the Centro-Parietal positively (CPP) is known to occur in tasks where a visual or auditory stimulus ramps up or down in signal intensity and RDM tasks (O'Connell et al., 2012; Kelly and O'Connell, 2013; Rangelov and Mattingley, 2020). Combining all these experimental considerations, the

best visual stimulus to test a simple hypothesis of the relationship of CPP to an evidence accumulation rate during decision-making might be a RDM where the participant must differentiate either leftward or rightward motion during each trial (i.e. a 2AFC task). Note that researchers should not be limited to prior research, and researchers should feel free to design new experimental tasks and new experimental designs. Diversity of experimental design ideas will always remain important for the growth of the field of joint modeling and the growth of knowledge.

4. Collection and preprocessing of M/EEG for joint modeling

4.1. Understanding artifactual processes in M/EEG data

M/EEG contains many overlapping sources of information, including brain-generated sources and artifactual sources (Nunez and Srinivasan, 2006; Nunez et al., 2016). Due to the complexity and the size of information in collected M/EEG, parameter estimation of joint models of behavior and M/EEG usually requires extraction of specific M/EEG signals. One important step in this process is the removal of artifact, specifically muscle-generated electromyographic (EMG) signals, environmental electrical artifacts, and physical movement artifacts that are prevalent in scalp-recorded EEG recordings (Whitham et al., 2007; Nunez et al., 2016). Unless those artifacts are removed, or specifically accounted for in a joint model, these sources of artifact will likely add noise that effects results of joint modeling. These artifacts could thus greatly influence the results of neurocognitive models in parameter estimates or the results of model comparisons (e.g. see Hawkins et al., 2017).

Electrical and movement artifacts in scalp-recorded EEG records can be reduced by proper recording practices. Usually the goal of these practices is to (1) keep consistent electrical contact of the ground and reference electrodes to the scalp, (2) keep consistent electrical contact of all other electrodes (or to remove those electrodes from later analysis, if some electrodes can be removed without much loss of information, such as in high-density EEG), and (3) by removing or shielding external sources of electrical artifact from the area of recording. Electrodes with higher impedances typically have larger amplitude environmental noise (with large noise amplitudes at 60 Hz or 50 Hz depending upon the place in the world in which the recording takes place) that can be removed with online or offline targeted filtering (Kappenman and Luck, 2010). But also large impedances can be indicative of an electrode that is not making consistent electrical contact with the scalp and could produce high amplitude movement artifact. Inconsistent contact of an electrode could lead to sudden changes in impedance that can affect EEG records of that electrode. Most commercially available EEG systems have easy-to-implement methods of measuring electrode impedance. Those electrodes that have very high impedances can be given special attention (using Ohm cutoffs depending upon the EEG cap and system), such as making sure that the electrode is making stable contact with the skin of the scalp through hair. Unless using “dry” EEG systems, an appropriate amount of conductive gel or saline is also important to maintain electrical contact. Although not using too much conductive solution is also important to avoid electrical bridging (Greischar et al., 2004). We recommend reading some of the existing recommendations for various EEG laboratories resources for discussions of best practices on EEG data collection (e.g. Farrens et al., 2020), and recent work on improving recording practices for different hair types (e.g. Etienne et al., 2020).

468 In addition to proper recording practices, additional artifact correction is almost always
469 performed of EEG data, especially for eye blink, EMG, cardiovascular-generated electrocar-
470 diographic (EKG) signals, and miscellaneous movement artifacts. For joint modeling this
471 artifact correction will usually be performed offline, as opposed to BCI applications where
472 artifact correction is performed online. Artifact correction is also helpful for magnetoen-
473 cephalographic (MEG) and intracranial electroencephalographic (iEEG) records, although
474 the prevalence of different types of artifacts will differ across modalities and recording sys-
475 tems. Techniques to mitigate artifacts before parameter estimation is especially necessary
476 due to the simple assumptions often made intrinsically in many joint models of M/EEG and
477 behavior. Because brain generated EEG recorded from the scalp will not manifest as sudden
478 large amplitude spikes after filtering by the skull and skin, common offline artifact correc-
479 tion techniques include removing epochs of data or specific electrode records that surpass a
480 particular amplitude cutoff are the most useful and basic artifact corrections.

481 Another popular method which deserves special consideration is Independent Component
482 Analysis (ICA Makeig et al., 1996) for scalp-recorded EEG and MEG. ICA finds linear
483 mixtures across M/EEG channels that are statistically independent as possible by finding
484 maximal non-Gaussian mixtures. The resulting components have M/EEG-like times courses
485 where, if the time dependency were ignored and the time samples were randomly shuffled, the
486 resulting distributions would have minimum mutual information or be separated on the basis
487 of kurtosis (Hyvärinen and Oja, 1997; Jung et al., 2000). In practice, these methods often
488 yield non-normal mixtures that have distributions with outliers, and thus ICA algorithms
489 are especially good at extracting certain artifacts within multiple EEG channels such as
490 eye blinks (IC1 in **Figure 3**), eye movements, EKG, and temporary changes in electrode
491 impedances (IC6 and IC7 in **Figure 3**). Sometimes ICA algorithms can also find some EMG
492 artifacts that can be easily identified in the EEG data (IC12 in **Figure 3**). These components
493 can then be extracted from multiple electrodes, and the resulting EEG data can be converted
494 back to channel space or be kept in component space for joint modeling. EEG components
495 (mixtures of electrodes) are discussed in detail below.

496 Note that it is unlikely that all artifact will be removed using these methods, especially
497 EMG artifact in scalp-recorded EEG (Whitham et al., 2007; Nunez et al., 2016). So to
498 further avoid artifact in scalp-EEG, it is best to choose specific methods of signal extraction.
499 This is typically performed outside of joint modeling, but future researchers may be able
500 to model these methods explicitly to retain sources of noise in the model. These methods
501 include popular EEG-band limited analyses (e.g. calculating 8 to 13 Hz alpha power over
502 posterior electrodes in Ghaderi-Kangavari et al., 2021) and event-related potential (ERP)
503 analyses that mitigate artifactual components through averaging across trials (e.g. N200
504 latencies in Nunez et al., 2019a).

505 Finally, M/EEG outliers could drive the the entire modeling results. A choice must be
506 made between explicit removal and explicit modeling of lapse and artifactual processes. Note
507 that there will be differences between artifactual (non-brain generated M/EEG) and brain
508 generated M/EEG that is related to *some* cognitive process but not related to the cognitive
509 process of interest. This could be because the participant is mind wandering during that
510 trial (e.g. see Hawkins et al., 2021). It is a question for new joint modeling work whether
511 inclusion of non-brain generated M/EEG artifactual processes and/or mixtures of cognitive
512 processes within the join modeling itself would be beneficial to understand the researchers'

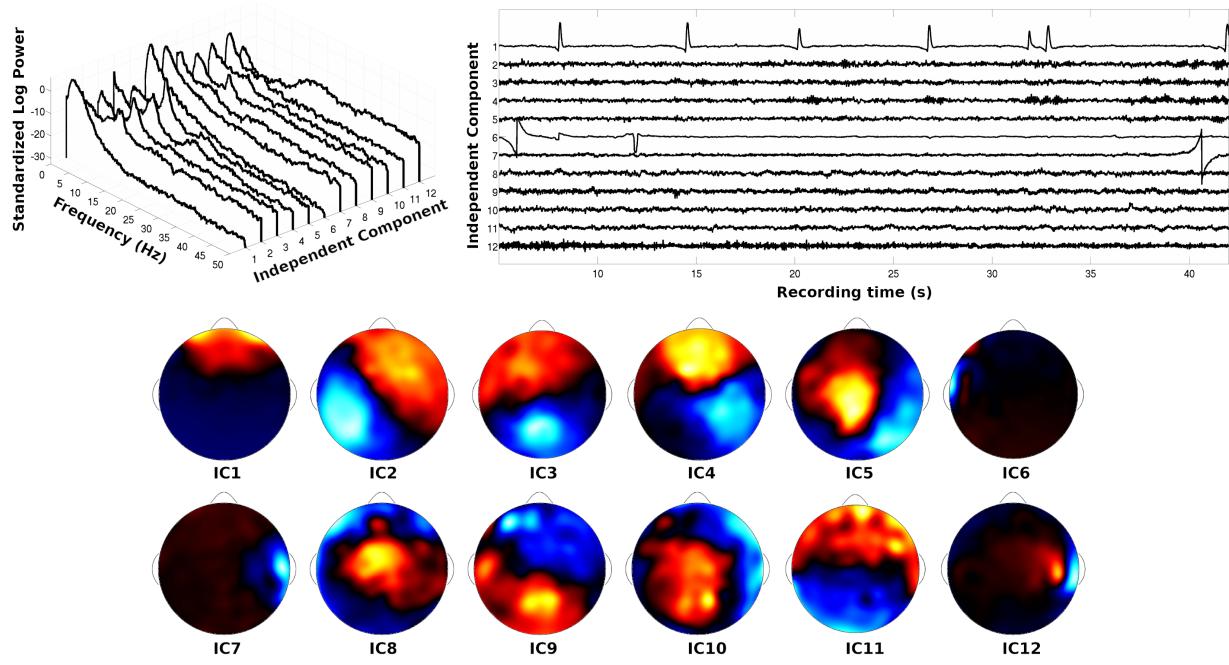


Figure 3: (From top-left, clockwise) Power spectra, time courses, and spline-interpolated channel weights from the first twelve Independent Components (ICs). The Independent Component Analysis (ICA) algorithm was performed on an EEG record in which a participant was fixating on a computer monitor. ICs that are likely to reflect artifact can be subtracted from the EEG data before neurocognitive modeling. IC1 is indicative of an eye blink. IC6 and IC7 are indicative of temporary changes in channel impedances. IC12 is indicative of some muscle artifact. Note this figure is adapted with permission from Figure 4 by Nunez et al. (2016)

513 specific hypotheses and questions.

514 *4.2. Extraction of relevant M/EEG signals*

515 In previous research, most researchers have extracted specific EEG signals before joint
516 modeling (e.g. Frank et al., 2015). Targeted extraction of specific EEG signals can be espe-
517 cially beneficial for joint modeling when those signals have a rich literature of prior research
518 and cognitive theory. We will concentrate here on popular EEG signals studied within Cog-
519 nitive Neuroscience.

520 Event Related Potentials (ERPs) are defined as averages of M/EEG across experimental
521 trials, time-locked to specific events such as the onset of a visual stimulus (a Visual Evoked
522 potential; VEP) or the execution of a response such as a button press (a Motor Evoked
523 Potential; MEP). ERPs can be found by a simple average of M/EEG v over N trials i such
524 that resulting signal μ varies over a time index t , time-locked to an experimental event:

$$\mu_t = \frac{1}{N} \sum_{i=1}^N v_{ti} \quad (10)$$

525 ERPs also have rich literature (Luck, 2012), from which best practices can be recom-
526 mended. This literature can also be used to generate new confirmatory and exploratory
527 hypotheses that could be answered with joint modeling methods. After calculation of ERPs
528 using the equation above, specific peak latencies, amplitudes, or deviation times from baseline
529 are typically extracted, either positive or negative peaks. Common ERPs are the negative
530 N200 peak approximately 200 ms after the onset of a visual stimulus in occipital and parietal
531 electrodes (sometimes labeled N1 for the first negative peak), recently thought to encode
532 the onset of evidence accumulation during decision-making (Nunez et al., 2019a). Another
533 common ERP is the positive P300 peak at least 300 ms after the onset of a visual stimulus;
534 this ERP is also called the Cento-parietal Positivity (CPP) during specific decision-making
535 tasks (Twomey et al., 2015), discussed extensively earlier and used in example **Models 2** and
536 **3**. Common MEPs are the Readiness Potential (RP) and the related Lateralized Readiness
537 Potential (LRP). ERPs can also be estimated on single-trials (e.g. Nunez et al., 2017, 2019a).

538 Frequency or time-frequency decompositions such as Fourier and wavelet analyses are
539 also common methods to extract specific signals in M/EEG. These decompositions also the
540 basis of derived measures such as M/EEG coherence. These typically rely on signal to noise
541 ratios and have been used in neurocognitive modeling of decision-making (Frank et al., 2015).
542 These can be event-locked or from endogenous rhythms not related to the timing of the task.
543 Many algorithms in high level programming languages such as the Fast-Fourier transform
544 are sufficient to estimate these signals. Although best practices in EEG and standard EEG
545 conventions should be known (Nunez et al., 2016). Alternatively there are some algorithms
546 developed specifically to extract specific band-limited waveforms, such as finding 80 – 250
547 Hz High Frequency Oscillations (HFOs) that last only for a few milliseconds in iEEG data
548 (Charupanit and Lopour, 2017; Nunez et al., 2022). One warning for scalp EEG is that high
549 frequencies (approx. > 20 Hz) typically contain more EMG artifact (Whitham et al., 2007),
550 and therefore care must be taken when interpreting the results of measures derived from
551 high frequencies embedded in joint models. Although MEG systems may be more robust

552 to EMG artifact (Claus et al., 2012; Muthukumaraswamy, 2013), and newer MEG systems
553 could be even more robust to EMG artifact (Ilmoniemi and Sarvas, 2019). Another caution is
554 that researchers have shown heterogeneity in the power bands across and within participants
555 (Nunez et al., 2001) as well as heterogeneity in the waveforms themselves (Donoghue et al.,
556 2021), so care must be taken to extract specific signals for joint modeling.

557 Steady-state evoked potentials (SSEPs), and in particular steady-state visual evoked po-
558 tentials (SSVEPs) and related steady state auditory potentials (SSAEPs), are a special case
559 of band-limited analysis where the frequency band of interest in the M/EEG results from a
560 processing stimuli at a certain presentation rates or “flicker” (Regan, 1977). For instance a
561 Gabor patch flickering at 15 Hz will result in a large, narrow-band, 15 Hz response (and often
562 harmonics of 15 Hz) in EEG. This is the result of the cortex receiving and processing signals
563 at this rate, which is expected of a linear system. Some researchers have found evidence
564 that endogenous EEG signals may also “entrain” to the stimulus frequencies (Srinivasan,
565 2004; Ding et al., 2006). SSEP analyses could be particularly useful for fitting joint models
566 because the amplitude or phase-locking across trials is thought to index within individual
567 and individual differences in attention (Ding et al., 2006). We have previously explored how
568 individual differences in attention as measured by SSVEPs affected cognitive components of
569 decision-making (Nunez et al., 2015).

570 Typically ERPs, SSEPs, power in different (time-)frequency bands, are in replicable scalp
571 and brain locations. These specific EEG signals also often have replicable cognitive inter-
572 pretations found in the Cognitive Neuroscience literature, e.g. see **Table 1**. However the
573 exact electrode/sensor locations will differ participant to participant, and could even change
574 within a participant due to changing artifactual sources and changing electrical contact of the
575 electrodes over the course of a long experiment. Furthermore, within the field of model-based
576 cognitive neuroscience modeling, we should make better use of overlapping information in
577 M/EEG data to better describe cognitive theory (see Bridwell et al., 2018). For intracranial
578 EEG (iEEG) spatial filters can also be useful in extracting relevant EEG features for joint
579 modeling (e.g. see Schawronkow and Voytek, 2021). Therefore, weight averages across chan-
580 nels should also be considered. We will call these set of analysis “component analyses” for
581 finding mixtures, typically linear mixtures, of M/EEG data that may better reflect the un-
582 derlying source components as reflected on the scalp or in intracranial electrode space (Parra
583 et al., 2005). Component analyses will extract a weighted average of electrodes/sensors in
584 order to improve the signal to noise ratio of mixtures. Previously mentioned Independent
585 Component Analysis (ICA) is one example algorithm, but one could also consider Principal
586 Component Analysis (PCA) (e.g. Nunez et al., 2017, 2019a), Canonical Correlation Analysis
587 (CCA) (e.g. van Vugt et al., 2012), and explicitly modeling mixtures over electrodes/sensors
588 in joint models.

589 Preprocessing of M/EEG may be theoretically undesirable since extraction of specific
590 signals often involves removing potentially useful information from M/EEG signals that could
591 be better accounted for with statistical models in joint modeling. In all our previously
592 published work (Nunez et al., 2015, 2017, 2019a; Lui et al., 2021), we extracted specific
593 EEG potentials before joint modeling. These methods are useful for testing specific theories
594 of visual attention, motor processing, or visual encoding, but not useful for understanding
595 parallel processes that occur during decision-making. The future of joint modeling techniques
596 should better account for full M/EEG data to improve prediction and hypothesis testing.

597 This can be performed by embedding mixture models of M/EEG signals, either based on
598 brain connectivity and neural network behavior (with a model of electric volume conduction
599 to the scalp, see paper by Nunez et al. (2019b)) or more non-parametric methods based on
600 mixtures of oscillating signals.

601 **5. Implementing model fitting procedures and estimating parameters**

602 Finding parameter estimates from a proposed model can be difficult. There are many
603 more restrictions on parameter fitting than model simulation due to difficulty in maximizing
604 likelihood spaces or sampling from posterior distributions. Many joint models will also not
605 be identifiable, as discussed below. However multiple free programs exist to help you fit joint
606 models of M/EEG and behavior. Most require some knowledge of a programming language
607 and sampling methods.

608 *5.1. Avoiding model complexity*

609 To test specific hypotheses or compare theories, a perfect explanation of all relevant
610 M/EEG signals and behavioral data is often unnecessary and could result in overfitting the
611 model (see Navarro, 2019). The degree of complexity needed in the fitted model(s) will
612 depend on the goals of the researcher. In DDMs, choosing between two discrete choices is
613 assumed to occur due to particular time-varying sampling of relative evidence. The rate of
614 sampling of evidence is assumed to change both during a single choice due to within-trial
615 variability in a random walk process, but also change across many similar choices, due to
616 trial-to-trial variability (Ratcliff et al., 2016). In an experiment with the participant making
617 a choice during every experimental trial, expected across-trial changes in the parameters are
618 often modeled with across-trial variability parameters. While it is expected that humans do
619 have variability in strategy, attention, and response cautiousness etc. across trials, across-trial
620 variability parameters of a full DDM model cannot be easily estimated with behavior alone
621 (Boehm et al., 2018). This problem of parameter estimation can be at least slightly improved
622 with the addition of EEG signals on single trials (Nunez et al., 2017; Hawkins et al., 2017).
623 Therefore when fitting behavioral data to joint models, we often make the choice to fit simple
624 DDM without any across-trial variability in the DDM parameters that is not described by
625 the single-trial EEG measures (Nunez et al., 2017, 2019a). Other joint modeling research has
626 often included across-trial variability in evidence accumulation rates but no other parameters
627 (e.g. Frank et al., 2015; Ghaderi-Kangavari et al., 2022). In past work we have purposely not
628 included some across-trial variability parameters because we have shown in simulations that
629 the question of interest about the M/EEG-cognitive relationship can be answered without
630 more complex models, knowing that the greater mathematical theory of decision-making
631 does have variability and that the data would better be described by more complicated
632 models. However if we were differentiating between hypotheses that are very similar in their
633 predictions, we might require to include trial-to-trial variability parameters or more precise
634 M/EEG correlates of evidence accumulation.

635 *5.2. Parameter recovery of simulated models*

636 Simulation is especially important for newer models and model fitting procedures that
637 are not widely used. Because jointly fitting neural data and human behavior is not widely

used within Cognitive Neuroscience and Psychology, the majority of model fitting procedures that readers may implement will fall into this category. Thus researchers interested in neurocognitive models should always simulate and refit new models to understand parameter recovery. The parameter estimates from new models and new model fitting procedures should always be verified by refitting data from simulations before modeling results can be trusted to be self-consistent, whether or not the model is reflective of reality or a specific hypothesis. For instance, model-fitting procedures could give realistic results, but not recover the same parameters when simulated. This means that the parameter estimates recovered from model-fitting will not reflect reality, even if the simplified model is completely true. In addition, some parameters of a model may recover, and therefore be relevant parameters to analyze in real data, while other parameters of a model will not recover. In **Figure 4** we show recovery of the 5 of 6 parameters from **Model 3** when using Markov Chain Monte Carlo (MCMC) sampling in JAGS (discussed below) with the jags-wiener plugin (Wabersich and Vandekerckhove, 2013). We fit a model that assumes there is no extra trial-to-trial variability not due to trial-to-trial variability in CPP (e.g. assuming $\eta = 0$ when fitting the model). The code for this simulation, parameter recovery, and plot is given in https://github.com/mdnunez/pyhddmjags/blob/master/simpleCPP_test.py as of March 2022.

5.3. Comparing models

Model comparison is useful when multiple theories could describe the data and the evaluation of hypotheses would depend on the theory assumed. Model comparison can also be used to evaluate competing theories directly. Therefore model comparison in joint cognitive modeling of M/EEG and human behavior is almost always beneficial. Model comparison can be based on multiple dimensions, but typically researchers are interested in the models that provide the most predictive and/or realistic accounts of the neural and behavior data. Model comparison could therefore be based on how well the model predicts data or how well models evaluate the hypothesis versus an alternative hypothesis. For example, if we were interested in the hypothesis that the CPP reflects evidence accumulation *exactly*, we could compare a model where the slope of the CPP describes the evidence accumulation rate itself, e.g. **Model 3**, to a more general model where the relationship between the CPP and evidence accumulation rate can be any value (see **Exercises**). If **Model 3** describes the data nearly as well as the more generalized model, then we have evidence for the hypothesis that the CPP reflects evidence accumulation exactly.

How well a model “*predicts*” data used to fit the model itself is often used as a measure of performance (Blohm et al., 2020). This is called “in-sample” prediction. However evaluating models based only on in-sample prediction can result in over-fitting the data. Over-fitting describes the situation in which additional in-sample prediction is gained through model complexity that does not relate to the underlying true generative process and results in worse “out-of-sample” prediction. Out-of-sample prediction refers to how well a model predicts data that it was not fit to. Out-of-sample could also refer, although this is not traditionally in the definition, how well it generalizes to similar data from other experiments (Lee et al., 2019; Vandekerckhove et al., 2019). Note that prediction of in-sample and out-of-sample data can be used to compare models that differentiate specific hypotheses, but it is often not necessarily to perfectly describe or predict EEG and/or behavioral data due to the presence

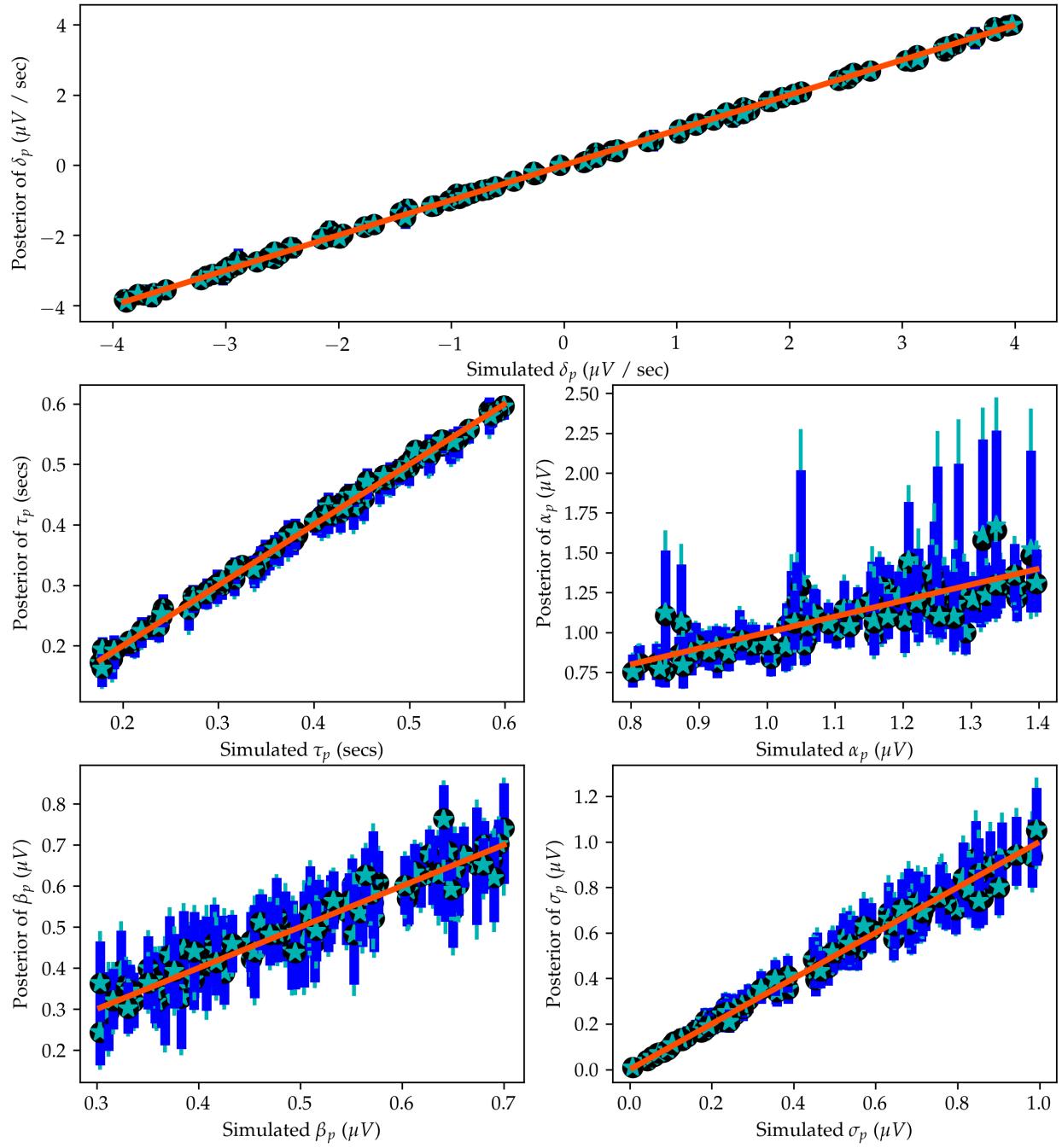


Figure 4: The recovery of 5 parameters from **Model 3**. We simulated **Model 3** using the provided Python code with 100 simulated participants and 100 trials for each participant. The x-axis of each plot is the true simulated parameter and the y-axis is a summary of posterior samples. The mean of the posterior distributions are given by teal star symbols and the median of the posterior distributions are given by black circles. Uncertainty about each parameter estimate is given by the 95% credible intervals of dark blue lines, and 99% credible intervals of teal lines.

682 of artifacts and noise in EEG and contaminate behavioral data not related to the cognition
683 of interest.

684 For evaluating neurocognitive models of M/EEG and human behavior, we prefer out-of-
685 sample prediction (e.g. Nunez et al., 2017; Schubert et al., 2019). Out-of-sample prediction
686 typically involves taking at least one subset of the data out before fitting the model to
687 the remaining “in-sample” data. For instance, one could split a data set where 80% of
688 the data is used to fit the model and 20% of the data is used to evaluate out-of-sample
689 prediction. One method to evaluate the similarity of predicted data to the actual data
690 is with a “proportion of variance explained” calculation. For instance, we have previously
691 calculated R_{pred}^2 of participants’ accuracy and correct reaction time 25th percentiles, medians,
692 and 75th percentiles (Nunez et al., 2015, 2017). R_{pred}^2 is a measure of percentage variance in a
693 statistic T (e.g. accuracy, correct-RT median, etc.) explained by in-sample or out-of-sample
694 prediction. It is a function of the mean squared error of prediction (MSEP) and the sample
695 variance of the statistic T based on a sample size J of data observations. R_{pred}^2 is defined as:

$$R_{\text{pred}}^2 = 1 - \frac{\sum_{j=1}^J (T_j - T_{(\text{pred})j})^2 / (J - 1)}{\sum_{j=1}^J (T_j - \bar{T})^2 / (J - 1)} = 1 - \frac{\text{MSEP}_T}{\widehat{\text{Var}[T]}} \quad (11)$$

696 As one example, we could compare the out-of-sample prediction of **Model 2A** to **Model**
697 **2B** that was fit to 80% of the data from **Experiment 1**. The observations for each statistic
698 T in the R_{pred}^2 equation would be for every participant j with sample size J being the number
699 of participants.

700 There are other methods to evaluate out-of-sample prediction, such as calculating the
701 log-likelihood under predicted data (e.g. see Figure 9 of Turner et al., 2016). Evaluation
702 of models’ prediction ability can also be evaluated with plots such as quantile-quantile (Q-
703 Q) plots of measured versus predicted data quantiles. Because we typically use Bayesian
704 methods, we generate posterior predictive distributions for in-sample and out-of-sample data.
705 We can then create posterior predictive coverage plots (e.g. see Supplementary Materials of
706 Nunez et al., 2017). Plotting often provides additional information to more quantifiable
707 measures such as R_{pred}^2 or a similar measure.

708 Cross-validation refers to methods where out-of-sample prediction is performed repeti-
709 tively on different subsets of data (e.g. a new 20% of the same data set iteratively). Because
710 cross-validation lowers the impact of outliers in the out-sample, cross validation can be useful
711 for modeling M/EEG data because the presence of outliers is commonplace. However care
712 must be taken to avoid changing parameters of the model or model fitting procedure based
713 on the results of cross validation because this would make the cross validation process less
714 reflective of the true predictive ability of the model.

715 If out-of-sample prediction is not available, then often penalizing by model complexity
716 after in-sample prediction is used. This is often why Information Criteria measures are
717 used (e.g. Ghaderi-Kangavari et al., 2021, 2022). Essentially these measures are in-sample
718 prediction measures that penalize for model complexity. Akaike Information Criterion (AIC),
719 Bayesian Information Criteria (BIC), Deviance Information Criteria (DIC), and re-weighted
720 variations of these measures are often used. However these measures may yield different
721 results. For instance, it is thought that BIC more often favors models that match the ground
722 truth while AIC more often favors models that are predictive of new data (Aho et al., 2014;

723 Chandrasekaran and Hawkins, 2019) Therefore, it is important to pick one ahead of time and
724 stick to it, or preregister the modeling analysis (Lee et al., 2019).

725 Simulation is also important if you wish to perform model comparison. Some models may
726 fit data better not because the underlying theory is a better reflection of reality, but because
727 the models capture some contaminant process better. That is, neither model is correct but
728 the worse-fitting model is a better description of reality. Simulations of multiple models with
729 contaminant processes before performing parameter recovery for each model can thus reveal
730 which model and model-fitting procedure best recover true parameters.

731 We separate here confirmatory research from exploratory research. When discovering
732 the influence of EEG measures on parameters that describe decision-making behavior, it is
733 beneficial to explore various model types that may better match the theoretical evidence
734 accumulation process. This can be achieved, for instance, by fitting parameters from differ-
735 ent sequential sampling models (SSMs) directly (preferred) or simulating a variety of SSMs
736 and then exploring how similar parameters can be recovered in other models. Confirmatory
737 research necessitates large samples, pre-deciding an analysis plan including the specific joint
738 model to test (and perhaps preregistration of that analysis plan and model), and requiring
739 strict standards for hypothesis acceptance such as in clinical trials (Lee et al., 2019). Al-
740 though there exists a spectrum between exploratory and confirmatory research that has been
741 discussed elsewhere (e.g. Devezer et al., 2021).

742 *5.4. Fitting complex models using Bayesian methods*

743 We generally prefer to use programs that use Bayesian Markov Chain Monte Carlo
744 (MCMC) sampling and allow a large amount of flexibility to change the model structure.
745 Programs such as JAGS (Plummer, 2003), Stan (Carpenter et al., 2017), and PyMC3/4
746 (Salvatier et al., 2016) all make this process incredibly easy by allowing you to write your
747 own complex models, but without needing to write your own samplers. Note that Bayesian
748 analysis can be quite easy to learn for those that have a background in some mathematics
749 and statistics. A nice introduction to Bayesian analysis is given by Etz and Vandekerckhove
750 (2018). See also books by McElreath (2020) and Gelman et al. (2014).

751 All the aforementioned programs also allow the modeler to easily implement hierarchical
752 parameters which can better account for variance across experimental conditions, partici-
753 pants, sessions, etc. Hierarchical parameters can often better account for data with multiple
754 modes of data (Lee, 2011; Turner et al., 2016), such as EEG and human choice response
755 times. Hierarchical models often yield better estimates of parameters due to “shrinkage”
756 towards the mean parameters rather than fitting a model per participant or experimental
757 condition, which could lead to overfitting and misestimation (see Chapter 5 of Gelman et al.,
758 2014). Examples for fitting behavioural DDMs and neurocognitive DDMs using Python,
759 JAGS, and Stan, with the models themselves written in JAGS and Stan code, are given in
760 the repository https://github.com/mdnunez/pyhddm_jags. We encourage readers to run
761 the example models in this repository if they are interested in using JAGS and Stan with
762 Python. Note that connectors to JAGS and Stan also exist in R and other programming
763 languages.

764 The programmatic implementation to generate parameter estimates from joint models
765 that we preferred in the past is JAGS (Plummer, 2003). JAGS is now a somewhat older
766 program, but nicely contains multiple MCMC samplers and chooses among them based on

767 the user-defined model. Custom distributions can also be added to JAGS (Wabersich and
768 Vandekerckhove, 2014). JAGS uses Bayesian MCMC samplers to fit models to data and can
769 easily fit joint models to multiple data types. For instance, we fit a simplified version of
770 **Model 3** in JAGS with 12,000 original samples in each of six chains for each parameter.
771 After removing the first 2,000 samples as a “warm-up” or “burn-in” and then keeping only
772 every 10th sample, i.e. using a “thinning” parameter of 10, this results in 1,000 posterior
773 samples in each chain for $1,000 * 6 = 6,000$ samples from the estimated posterior distributions
774 for each parameter.

775 To assess model whether the model is reaching a unique solution (i.e. unique joint pos-
776 terior distributions), we can both inspect our MCMC chains but also gauge certain model
777 convergence diagnostics (Gelman et al., 2014). The Gelman-Rubin statistic and the number
778 of effective samples are calculated (Gelman et al., 2014). The Gelman-Rubin statistic as-
779 sseses the convergence of MCMC samplers by comparing the between-chain variance to the
780 within-chain variance of each parameter, with Gelman-Rubin statistic ≤ 1.1 thought to be a
781 necessity for convergence. We also implemented the recommendation by Gelman et al. (2014)
782 (see footnote in the 3rd Edition on page 283) to split the chains in half before calculating
783 the Gelman-Rubin statistic in order to account for non-stationary chains. The “effective
784 number of samples” equation scales the total sample number for each parameter posterior by
785 autocorrelation in the chains in order to estimate an independent number of samples. Larger
786 effective samples for each parameter in the model are better. The chains for parameters
787 with the largest Gelman-Rubin statistics and smallest effective number of samples are also
788 visually inspected to ensure convergence. In publications we typically report the maximum
789 Gelman-Rubin statistics across all parameters, and we have recently started to report the
790 minimum number of effective samples across all parameters.

791 *5.5. Prior distributions in Bayesian models*

792 When using Bayesian methods we often must choose prior distributions of parameters,
793 this is also true for estimating joint models of M/EEG and behavior with Bayesian methods.
794 When possible, we pick prior distributions based on previous publications, and such that the
795 prior distributions have weight over plausible values of the parameters. For instance, random
796 draws from a normal distribution with a mean of .5 and a standard deviation of .25 will
797 result in 68.2% of those draws within .25 and .75, and 95.4% of those draws within 0 and
798 1. [0, 1] is the domain of the relative start point parameter β that encodes initial evidence
799 bias in a DDM. Therefore a normal distribution with mean .5 and standard deviation of
800 .25 truncated to the domain [0, 1] would be a good prior distribution for this parameter,
801 disregarding algorithmic reasons why we might pick different priors (such as in true Gibbs
802 sampling).

803 There is ongoing research about what the best prior distributions are, depending upon
804 the type of sampler. There are also many philosophical discussions about whether to use
805 “informative” (generally narrow) or “weakly informative” (generally wide) priors. As mod-
806 elers we should experiment with different priors in simulation to see how and if they change
807 the results significantly. However we have not found that different “weakly informative”
808 priors changed results much based on posterior distributions of hierarchical DDM param-
809 eters. Prior distributions can change posterior distributions if those priors are very nar-
810 row, such that values in that parameter’s domain are near impossible (“informative” pri-

ors). For instance, a prior of $\beta \sim Normal(.5, .01^2)$ would restrict the posterior distribution of β to be approximately in the domain [.46, .54], within 4 standard deviations on both sides of the mean. We refer readers to a discussion about this topic by Andrew Gelman at <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>

5.6. Assessing posterior distributions

Posterior distributions provide evidence for parameters given the data and specific model architecture. They are influenced by prior distributions, but are often much more influenced by the data itself. Because the definition of probability in Bayesian analysis is uncertainty, we can inspect the posterior distributions themselves in order to calculate the probability of observing certain values of a parameter given the specific model and data. For instance, the posterior distributions that result from fitting **Model 2B** to data from **Experiment 2** would result in posterior distributions for all parameters, including both effect parameters ξ_{1j1} for experimental condition (1) and ξ_{1j2} for experimental condition (2). We could calculate the probability that each effect parameter is greater than 0.5 (for instance) by finding the proportion of posterior samples that are above 0.5, thus approximating the area under the curves of the posterior distributions, and thus approximating the probability (e.g. evidence) that each effect parameter is greater than 0.5. We can also calculate posterior distributions to answer other questions using a transformation of our parameters. For instance, we can calculate the posterior distribution of the difference between these two effects by matching MCMC samples of the original model fit to get one difference posterior of the quantity $\xi_{1j1} - \xi_{1j2}$ for participant j . We can then calculate the probability that $\xi_{1j1} > \xi_{1j2}$ by finding the proportion of posterior samples of the new quantity $\xi_{1j1} - \xi_{1j2}$ that are above zero.

Bayes Factors (BFs) usually provide the degree of evidence (defined as a probability ratio in Bayesian statistics) of data given a model where the parameter value λ is exactly some value x ($\lambda = x$) versus the same model where the parameter λ can take any other realistic value ($\lambda \neq x$) (Jeffreys, 1961; Kass and Raftery, 1995; Rouder and Morey, 2012; van Ravenzwaaij and Etz, 2021). BFs can also be inverted to give evidence for the more general case ($\lambda \neq x$) compared to the more specific case ($\lambda = x$). Generally Bayes Factors over 3 are considered positive evidence for the numerator model over the denominator model (e.g. the effect is 3 times more likely under the specific case than the general model) while over 20 is strong evidence (Kass and Raftery, 1995). Bayes Factors for parameters of joint models estimated with Bayesian methods can often be estimated using the Savage-Dickey density ratio of the posterior density of parameter λ at test value x over the prior density of parameter λ at test value x (Dickey and Lientz, 1970; Verdinelli and Wasserman, 1995; Wagenmakers et al., 2010; van Ravenzwaaij and Etz, 2021). For instance we have previously estimated specific Bayes Factors (BF1s) of linear relationships between non-decision time τ parameters and N200 ERP latencies that describe the amount of relative evidence of the effect parameter λ being equal to 1 ($\lambda = 1$) to a more general comparison model ($\lambda \neq 1$) using the Savage-Dickey density ratio (Nunez et al., 2019a). These BF1s compared the hypothesis of a “spike” distribution at 1 with no uncertainty in possible effect values ($\lambda = 1$) versus a model with less specific effect values ($\lambda \neq 1$). Note that Bayes Factors can describe comparisons of models generally (Etz and Vandekerckhove, 2018; van Doorn et al., 2021). However for joint modeling purposes we are often interested in comparing point hypotheses to general cases. Other BFs to compare complex models are currently difficult to calculate, and therefore we

855 recommend model comparisons using the methods previously mentioned (e.g. R^2_{pred}) for cases
856 other than comparing point hypotheses to general models.

857 **6. Discussion**

858 *6.1. Modeling of M/EEG generators*

859 Note that in this paper we assumed summary measures of M/EEG. We could instead
860 model and simulate M/EEG time-series, waveforms, or frequency bands for multiple elec-
861 trodes/sensors. Considerable work has been conducted to understand the neural generators
862 of M/EEG (e.g. David and Friston, 2003; Nunez and Srinivasan, 2006; Srinivasan et al., 2013),
863 especially in the field of Computational Neuroscience. To this end, many researchers in the
864 field of Computational Neuroscience have built models of M/EEG, scaled up from single-
865 unit neuron activity, the activity of populations of neurons, and the connectivity between
866 neural populations (Daunizeau et al., 2011; Glomb et al., 2021). Developing and simulating
867 these models have been helpful in generating theory to understand how the brain generates
868 M/EEG, and there has been some success relating model parameters to human behavior.

869 However finding parameter estimates of these models often suffers from the *inverse prob-*
870 *lem*, which is known in statistical modeling as *unidentifiability*. The inverse problem is a
871 concept to be aware of across all modeling in all scientific fields. Inverse problems / unidenti-
872 fiabilities occur when there is more than one unique parameter set that will describe the data
873 when the generator model is known. Therefore, we cannot *invert*, or *find unique parameter*
874 *estimates*, of a model even though the model can be simulated and we have a good idea of the
875 theoretical concepts. In M/EEG computational modeling, this is often due to the fact that
876 the M/EEG *sources* are typically not constrained to a two-dimensional representation of the
877 brain, but instead a three-dimensional representation of the brain, while the scalp electrodes
878 (EEG), MEG sensors, or intracranial electrodes (iEEG) exist on curved two-dimensional
879 surfaces. Specifically the two-dimensional surfaces are usually the scalp, MEG helmet, or
880 electrode strips respectively. Finding these EEG *sources* at specific locations in the brain
881 require further assumptions about the EEG generators to be made (e.g. prior information in
882 Bayesian models, see Cai et al., 2018). Unfortunately these assumptions have been difficult
883 to prove experimentally due to the invasiveness of surgical procedures, the nature of electrical
884 volume conduction in the head, and the complexity of the brain. And many studies testing
885 inverse problem solutions for M/EEG rely on comparison to other algorithms. Although this
886 may change with the advent of new MEG technology (Ilmoniemi and Sarvas, 2019). Thus
887 some researchers rely on finding only *representative* sources in M/EEG (Nunez et al., 2019b),
888 including in our own work (e.g. Nunez et al., 2019a).

889 M/EEG records are an extremely rich data source. We usually assume many M/EEG
890 sources in the brain, e.g. dipole sheet, that generate time series data of multiple electrodes,
891 usually 16-256 electrodes in scalp EEG, approximately 300 sensors in MEG, and often > 50
892 electrodes in intracranial EEG, that are typically sampled at least 250 Hz. This results in
893 rich multivariate time series data with some spatial information. There are specific M/EEG
894 waveforms, power bands, network interactions, etc. that exist within the data. But there is no
895 reason to suspect that all useful measures within M/EEG have been studied, and we should
896 expect that there are measures of M/EEG that can predict behavior and cognition that have
897 yet to be found. Even with the advent of machine learning techniques which can explore rich

898 data sets, we expect neurocognitive modeling to reveal more about M/EEG data in the future.
899 Within M/EEG multivariate time series, there are also artifactual components, discussed
900 previously, embedded in both scalp EEG, MEG, and intracranial EEG, but especially scalp
901 EEG and older MEG systems. Thus, including all relevant information for modeling full
902 time-series of M/EEG in a theoretical model can be difficult.

903 However we do not need to find source estimates of M/EEG to better understand the neu-
904 rocognitive theory. We could, for instance, translate specific EEG potentials into cognition,
905 EEG potentials like the CPP slope, which in **Model 3** is hypothesized to be generated from
906 mean evidence accumulation rate during a trial. We could also instead build models that
907 describe the M/EEG phenomena, such as specific waveforms, while not assuming any partic-
908 ular type of cognition or brain activity. That is, we could describe the observed phenomena
909 and not develop a neurocognitive understanding of M/EEG. This may aid us by allowing
910 measurement noise in observed M/EEG potentials in our models, for instance. One promis-
911 ing method is to simulate EEG time courses by simulating from Morlet wavelet transforms
912 (Bridwell et al., 2018). Specific noise in M/EEG could also be simulated to understand the
913 robustness of the modeling procedure. If specific signals from M/EEG are extracted for joint
914 modeling, such as ERPs or power from a certain frequency band, capturing a full possible
915 range of these signals and contaminants in simulation is beneficial to test the robustness of
916 the procedure (Hawkins et al., 2017).

917 6.2. The future of joint modeling

918 We do not consider our past work of modeling EEG and human behavior to be “true”
919 joint modeling. We used hierarchical Bayesian methods, but only assumed simple linear influ-
920 ences of EEG measures on cognitive parameters, similar to **Model 2** presented above. These
921 methods could be consider simple “directed” approaches described by Palestro et al. (2018).
922 Ultimately, as Cognitive Neuroscientists and Cognitive Modelers, we would like to develop
923 computational theory that predicts both observed human behavior and EEG dynamics, such
924 as in **Model 3**. Among the several approaches described by Turner et al. (2017); Palestro
925 et al. (2018), researchers should ultimately seek to use an *integrative* approach with *simulta-*
926 *neous modeling* of EEG and behavior to test neurocognitive theory. In this way of thinking
927 about model-based cognitive neuroscience, what we have typically performed is *simultaneous*
928 *joint modeling* with linear connectors between EEG measures and decision-making behav-
929 ior, but not using an *integrative* approach (Nunez et al., 2015, 2017, 2019a). Furthermore
930 researchers, including ourselves, conducting joint modeling studies have typically fit simulta-
931 neous joint models with linear connectors between EEG measures and cognitive parameters
932 (e.g. Frank et al., 2015; Nunez et al., 2017; van Ravenzwaaij et al., 2017; Schubert et al.,
933 2019). Future research should improve upon previous joint modeling work. In the future
934 we wish to use the richness of EEG data and more informative human behavioral measures
935 (e.g. eye-tracking) within joint modeling frameworks to answer important neurocognitive
936 questions. This work will also lead to better *integrative* joint models with possible non-linear
937 connections.

938 Recently developed model fitting procedures are making fitting joint models easier. We
939 are particularly excited about algorithms that allow sampling from posteriors of joint models
940 when a likelihood is not available in closed-form or difficult to derive and estimate. One
941 particular promising program is BayesFlow, which finds posterior samples from simulation-

942 based models using invertible neural networks (Radev et al., 2020; Schmitt et al., 2022). A
943 similar promising method is to use neural networks to learn approximate likelihoods that can
944 then be used to find posterior distributions of joint models (Fengler et al., 2021). In general
945 we expect future model fitting procedures to be more flexible in the types of models that can
946 be fit to data, making joint modeling of M/EEG and behavioral data easier to implement.

947 *6.3. Conclusion*

948 We hope this tutorial serves as a guide for those researchers and students interested in
949 joint modeling of M/EEG and behavior. We have covered the possible motivations to per-
950 form joint modeling, the definition of joint models, building of joint models, simulating joint
951 models, experimental design, artifactual processes in M/EEG data, specific M/EEG signals,
952 model fitting implementations, parameter recovery, model comparison, and the future of
953 joint modeling. We have focused our examples on the relationship of scalp-recorded EEG
954 and decision-making. In particular we have used a guiding example of testing the hypoth-
955 esized relationship of the Centro-Parietal Positivity (CPP) slope to evidence accumulation
956 rate. However these techniques and principles can easily be applied to other neurocognitive
957 domains and questions using both animal and human electrophysiology. We expect joint
958 modeling to be able to answer questions that cannot be answered with other methods be-
959 cause joint modeling allows direct testing of neurocognitive theory. And we look forward to
960 reading about future research using joint modeling of M/EEG and behavior.

961 **7. Exercises**

- 962 1. Run the Python code from Model 3 or rewrite the code in another language (e.g. R)
963 and then run the code. Plot histograms or density approximations of the response time,
964 accuracy, and CPP slope data for some participants.
- 965 2. What model could be compared to Model 3 in order to test the hypothesis that the
966 CPP slope on each trial is a reflection of evidence accumulation? Specifically, what
967 model along with Model 3 could be fit to CPP slopes, response times, and accuracies
968 to test this hypothesis?
- 969 3. How could we change Model 3 to test the hypothesis that the CPP slope reflects a *scaled*
970 version of evidence accumulation rate, that is the mean rate of evidence accumulation
971 δ is not in micro-volts μV . Assume that the CPP slope could be scaled differently in
972 each participant due to scalp volume conduction differences across participants.
- 973 4. Using the given Python simulation of Model 3 as a guide, simulate from Model 1 while
974 assuming that the CPP slopes come from a distribution of $Normal(3, 1^2)$ μV (micro-
975 volts) per second across trials, that θ_0 is equal to .25 seconds, θ_1 is equal to .1, σ is
976 equal to .1, γ_0 is equal to 0, and γ_1 is equal to .3. Does this model produce similar
977 (and in this case of Model 3, more realistic) values of response times r and accuracies
978 x across trials n compared to Model 3?

979 Solutions are located after the **References** section of this document.

980 **Acknowledgements**

981 Special thanks to Kitty K. Lui, Keith Barrett, Kiana Scambray, Aishwarya K. Gosai,
982 Michelle Cheung, and Josh Tromberg for help with EEG collection and artifact removal. Cort
983 Horton is well appreciated for creating useful MATLAB functions for EEG visualization and
984 artifact removal. Aaron Bornstein is thanked for his help on similar table to **Table 1** that
985 included fMRI measures. Special thanks to Konrad Mikalauskas for proofreading a version
986 of this paper.

987 This research was supported by grants #1658303, #1850849, and #2051186 from the
988 United States (US) National Science Foundation (NSF), US National Institutes of Health
989 (NIH) grant 2R01MH68004 to RS, and US NSF grants to JV #1230118 and #1534472.

990 **Conflict of interest statement**

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

993 **Open practices statement**

994 All the code and data simulations from this work are given in the following repository
995 as of March 2022: <https://github.com/mdnunez/pyhddmjags>. These analyses were not
996 preregistered.

997 **References**

- 998 Aho, K., Derryberry, D., and Peterson, T. (2014). Model selection for ecologists: the world-
999 views of AIC and BIC. *Ecology*, 95(3):631–636. Publisher: Wiley.
- 1000 Au, J., Katz, B., Moon, A., Talati, S., Abagis, T. R., Jonides, J., and Jaeggi, S. M.
1001 (2021). Post-training stimulation of the right dorsolateral prefrontal cortex impairs work-
1002 ing memory training performance. *Journal of Neuroscience Research*, n/a(n/a). eprint:
1003 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jnr.24784>.
- 1004 Blohm, G., Kording, K. P., and Schrater, P. R. (2020). A How-to-Model Guide for Neuro-
1005 science. *eNeuro*, 7(1).
- 1006 Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L., and Stahl, J. (2012). Pre-
1007 dicting Perceptual Decision Biases from Early Brain Activity. *Journal of Neuroscience*,
1008 32(36):12488–12498.
- 1009 Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Kryptotos,
1010 A.-M., Lerche, V., Logan, G. D., Palmeri, T. J., et al. (2018). Estimating across-trial vari-
1011 ability parameters of the diffusion decision model: Expert advice and recommendations.
1012 *Journal of Mathematical Psychology*, 87:46–75.
- 1013 Boehm, U., van Maanen, L., Forstmann, B., and van Rijn, H. (2014). Trial-by-trial fluctua-
1014 tions in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage*,
1015 96:95–105.

- 1016 Bridwell, D. A., Cavanagh, J. F., Collins, A. G. E., Nunez, M. D., Srinivasan, R., Stober,
1017 S., and Calhoun, V. D. (2018). Moving Beyond ERP Components: A Selective Review of
1018 Approaches to Integrate EEG and Behavior. *Frontiers in Human Neuroscience*, 12:106.
- 1019 Cai, C., Sekihara, K., and Nagarajan, S. S. (2018). Hierarchical multiscale Bayesian algorithm
1020 for robust MEG/EEG source reconstruction. *NeuroImage*, 183:698–715.
- 1021 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M.,
1022 Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Program-
1023 ming Language. *Journal of Statistical Software*, 76(1):1–32. Number: 1.
- 1024 Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J.,
1025 and Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence
1026 over decision threshold. *Nature Neuroscience*, 14(11):1462–1467.
- 1027 Chandrasekaran, C. and Hawkins, G. E. (2019). ChaRTr: An R toolbox for modeling choices
1028 and response times in decision-making tasks. *Journal of Neuroscience Methods*, 328:108432.
- 1029 Charupanit, K. and Lopour, B. (2017). A simple statistical method for the automatic detec-
1030 tion of ripples in human intracranial EEG. *Brain topography*, 30(6):724–738.
- 1031 Claus, S., Velis, D., Lopes da Silva, F. H., Viergever, M. A., and Kalitzin, S. (2012). High
1032 frequency spectral components after Secobarbital: The contribution of muscular origin—A
1033 study with MEG/EEG. *Epilepsy Research*, 100(1):132–141.
- 1034 Crone, N. E., Miglioretti, D. L., Gordon, B., Sieracki, J. M., Wilson, M. T., Uematsu,
1035 S., and Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with
1036 electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization.
1037 *Brain*, 121(12):2271–2299.
- 1038 Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling: A critical
1039 review of the biophysical and statistical foundations. *NeuroImage*, 58(2):312–322.
- 1040 David, O. and Friston, K. J. (2003). A neural mass model for MEG/EEG:: coupling and
1041 neuronal dynamics. *NeuroImage*, 20(3):1743–1755.
- 1042 Devezer, B., Navarro, D. J., Vandekerckhove, J., and Ozge Buzbas, E. (2021). The case
1043 for formal methodology in scientific reform. *Royal Society Open Science*, 8(3):200805.
1044 Publisher: Royal Society.
- 1045 Dickey, J. M. and Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about
1046 chances, the order of a markov chain. *The Annals of Mathematical Statistics*, 41(1):214–
1047 226.
- 1048 Ding, J., Sperling, G., and Srinivasan, R. (2006). Attentional modulation of SSVEP power
1049 depends on the network tagged by the flicker frequency. *Cerebral cortex (New York, N.Y.
1050 : 1991)*, 16(7):1016–1029.

- 1051 Donoghue, T., Schawronkow, N., and Voytek, B. (2021). Methodological considerations
1052 for studying neural oscillations. *European Journal of Neuroscience*, n/a(n/a). eprint:
1053 <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.15361>.
- 1054 Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., Hawkins,
1055 G. E., Heathcote, A., Holmes, W. R., Kryptos, A.-M., Kupitz, C. N., Leite, F. P., Lerche,
1056 V., Lin, Y.-S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., van Maanen,
1057 L., van Ravenzwaaij, D., Vandekerckhove, J., Visser, I., Voss, A., White, C. N., Wiecki,
1058 T. V., Rieskamp, J., and Donkin, C. (2019). The Quality of Response Time Data Inference:
1059 A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic
1060 Bulletin & Review*, 26(4):1051–1069.
- 1061 Etienne, A., Laroia, T., Weigle, H., Afelin, A., Kelly, S. K., Krishnan, A., and Grover, P.
1062 (2020). Novel Electrodes for Reliable EEG Recordings on Coarse and Curly Hair. *bioRxiv*,
1063 page 2020.02.26.965202. Publisher: Cold Spring Harbor Laboratory Section: New Results.
- 1064 Etz, A. and Vandekerckhove, J. (2018). Introduction to Bayesian Inference for Psychology.
1065 *Psychonomic Bulletin & Review*, 25(1):5–34.
- 1066 Farrens, J., Simmons, A., Luck, S., and Kappenman, E. (2020). Electroencephalogram (EEG)
1067 Recording Protocol for Cognitive and Affective Human Neuroscience Researc. Technical
1068 report. Type: article.
- 1069 Fengler, A., Govindarajan, L. N., Chen, T., and Frank, M. J. (2021). Likelihood approxi-
1070 mation networks (LANs) for fast inference of simulation models in cognitive neuroscience.
1071 *eLife*, 10:e65074. Publisher: eLife Sciences Publications, Ltd.
- 1072 Forstmann, B., Ratcliff, R., and Wagenmakers, E.-J. (2016). Sequential Sampling Models
1073 in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of
1074 Psychology*, 67(1):641–666.
- 1075 Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., and
1076 Badre, D. (2015). fmri and eeg predictors of dynamic decision parameters during human
1077 reinforcement learning. *Journal of Neuroscience*, 35(2):485–494.
- 1078 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014).
1079 *Bayesian Data Analysis*. Taylor & Francis Group, LLC, Boca Raton, FL, 3rd edition.
- 1080 Ghaderi-Kangavari, A., Parand, K., Ebrahimpour, R., Nunez, M. D., and Rad, J. A. (2021).
1081 How spatial attention affects the decision process: looking through the lens of Bayesian
1082 hierarchical diffusion model & EEG analysis. Technical report. Company: Cold
1083 Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring
1084 Harbor Laboratory Section: New Results Type: article.
- 1085 Ghaderi-Kangavari, A., Rad, J. A., Parand, K., and Nunez, M. D. (2022). Neuro-cognitive
1086 models of single-trial eeg measures describe latent effects of spatial attention during per-
1087 ceptual decision making. *bioRxiv*.

- 1088 Gherman, S. and Philiastides, M. G. (2018). Human VMPFC encodes early signatures of
1089 confidence in perceptual decisions. *eLife*, 7:e38293. Publisher: eLife Sciences Publications,
1090 Ltd.
- 1091 Glomb, K., Cabral, J., Cattani, A., Mazzoni, A., Raj, A., and Franceschiello, B. (2021).
1092 Computational Models in Electroencephalography. *Brain Topography*.
- 1093 Gluth, S., Rieskamp, J., and Büchel, C. (2013). Classic EEG motor potentials track the
1094 emergence of value-based decisions. *NeuroImage*, 79:394–403.
- 1095 Greischar, L. L., Burghy, C. A., van Reekum, C. M., Jackson, D. C., Pizzagalli, D. A.,
1096 Mueller, C., and Davidson, R. J. (2004). Effects of electrode density and electrolyte spread-
1097 ing in dense array electroencephalographic recording. *Clinical Neurophysiology*, 115(3):710–
1098 720.
- 1099 Guest, O. and Martin, A. E. (2021). How Computational Modeling Can Force Theory
1100 Building in Psychological Science. *Perspectives on Psychological Science*, 16(4):789–802.
1101 Publisher: SAGE Publications Inc.
- 1102 Hautus, M. J., Macmillan, N. A., and Creelman, C. D. (2021). *Detection theory: A user's
1103 guide*. Routledge.
- 1104 Hawkins, G., Mittner, M., Forstmann, B., and Heathcote, A. (2021). Self-reported mind
1105 wandering reflects executive control and selective attention. *PsyArXiv*.
- 1106 Hawkins, G. E., Cavanagh, J. F., Brown, D., S., and Steyvers, M. (2022). Cognitive models
1107 as a tool to link decision behavior with eeg signals. In Turner, B. M. and Forstmann,
1108 B. U., editors, *Model Based Cognitive Neuroscience*. Springer New York, New York, NY.
1109 Forthcoming.
- 1110 Hawkins, G. E., Mittner, M., Forstmann, B. U., and Heathcote, A. (2017). On the effi-
1111 ciency of neurally-informed cognitive models to identify latent cognitive states. *Journal of
1112 Mathematical Psychology*, 76:142–155.
- 1113 Hyvärinen, A. and Oja, E. (1997). A Fast Fixed-Point Algorithm for Independent Component
1114 Analysis. *Neural Computation*, 9(7):1483–1492.
- 1115 Ilmoniemi, R. J. and Sarvas, J. (2019). *Brain Signals: Physics and Mathematics of MEG
1116 and EEG*. MIT Press. Google-Books-ID: wZiWDwAAQBAJ.
- 1117 Jagannathan, S. R., Bareham, C. A., and Bekinschtein, T. A. (2021). Decreasing alertness
1118 modulates perceptual decision-making. *Journal of Neuroscience*. Publisher: Society for
1119 Neuroscience Section: Research Articles.
- 1120 Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.
- 1121 Jensen, O. and Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha
1122 Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, 4.

- 1123 Jun, E. J., Bautista, A. R., Nunez, M. D., Allen, D. C., Tak, J. H., Alvarez, E., and Basso,
1124 M. A. (2021). Causal role for the primate superior colliculus in the computation of evidence
1125 for perceptual decisions. *Nature Neuroscience*, pages 1–11.
- 1126 Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Irugui, V., and Se-
1127 jnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separa-
1128 tion. *Psychophysiology*, 37(02):163–178.
- 1129 Kappenman, E. S. and Luck, S. J. (2010). The effects of electrode impedance on data quality
1130 and statistical significance in ERP recordings. *Psychophysiology*, 47(5):888–904. eprint:
1131 <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2010.01009.x>.
- 1132 Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical
1133 Association*, 90(430):773–795.
- 1134 Kelly, S. P. and O’Connell, R. G. (2013). Internal and External Influences on the Rate of Sen-
1135 sory Evidence Accumulation in the Human Brain. *Journal of Neuroscience*, 33(50):19434–
1136 19441.
- 1137 Klatt, L.-I., Schneider, D., Schubert, A.-L., Hanenberg, C., Lewald, J., Wascher, E., and
1138 Getzmann, S. (2020). Unraveling the Relation between EEG Correlates of Attentional
1139 Orienting and Sound Localization Performance: A Diffusion Model Approach. *Journal of
1140 Cognitive Neuroscience*, 32(5):945–962.
- 1141 Kohl, C., Spieser, L., Forster, B., Bestmann, S., and Yarrow, K. (2020). Centroparietal activ-
1142 ity mirrors the decision variable when tracking biased and time-varying sensory evidence.
1143 *Cognitive Psychology*, 122:101321.
- 1144 Kording, K., Blohm, G., Schrater, P., and Kay, K. (2018). Appreciating diversity of goals in
1145 computational neuroscience. *PsyArXiv*.
- 1146 Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models.
1147 *Journal of Mathematical Psychology*, 55(1):1–7.
- 1148 Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder,
1149 J. N., Trueblood, J. S., White, C. N., and Vandekerckhove, J. (2019). Robust Modeling in
1150 Cognitive Science. *Computational Brain & Behavior*, 2(3):141–153.
- 1151 Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., and O’Connell,
1152 R. G. (2016). Target Selection Signals Influence Perceptual Decisions by Modulating the
1153 Onset and Rate of Evidence Accumulation. *Current Biology*, 26(4):496–502.
- 1154 Luck, S. J. (2012). Event-related potentials. In *APA handbook of research methods in psy-
1155 chology, Vol 1: Foundations, planning, measures, and psychometrics*, APA handbooks in
1156 psychology®, pages 523–546. American Psychological Association, Washington, DC, US.
- 1157 Lui, K. K., Nunez, M. D., Cassidy, J. M., Vandekerckhove, J., Cramer, S. C., and Srinivasan,
1158 R. (2021). Timing of readiness potentials reflect a decision-making process in the human
1159 brain. *Computational Brain & Behavior*, 4(3):264–283.

- 1160 Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component
1161 analysis of electroencephalographic data. *Advances in neural information processing sys-*
1162 *tems*, pages 145–151.
- 1163 Manning, C., Hassall, C. D., Laurence, T. H., Norcia, A. M., Wagenmakers, E.-J., Snowling,
1164 M. J., Scerif, G., and Evans, N. J. (2021). Visual motion and decision-making in dyslexia:
1165 Evidence of reduced accumulation of sensory evidence and related neural dynamics. Tech-
1166 nical report. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring
1167 Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Type: article.
- 1168 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*
1169 *Stan*. Chapman and Hall/CRC, New York, 2 edition.
- 1170 McFarland, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and Beta
1171 Rhythm Topographies During Motor Imagery and Actual Movements. *Brain Topography*,
1172 12(3):177–186.
- 1173 Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in
1174 MEG/EEG: A review and recommendations. *Frontiers in Human Neuroscience*, 7.
- 1175 Navarro, D. J. (2019). Between the Devil and the Deep Blue Sea: Tensions Between Scientific
1176 Judgement and Statistical Model Selection. *Computational Brain & Behavior*, 2(1):28–34.
- 1177 Newsome, W. T. and Pare, E. B. (1988). A selective impairment of motion perception follow-
1178 ing lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8(6):2201–
1179 2211. Publisher: Society for Neuroscience Section: Articles.
- 1180 Nunez, M. D., Charupanit, K., Sen-Gupta, I., Lopour, B. A., and Lin, J. J. (2022). Beyond
1181 rates: time-varying dynamics of high frequency oscillations as a biomarker of the seizure
1182 onset zone. *Journal of Neural Engineering*, 19(1):016034. Publisher: IOP Publishing.
- 1183 Nunez, M. D., Gosai, A., Vandekerckhove, J., and Srinivasan, R. (2019a). The latency of a
1184 visual evoked potential tracks the onset of decision making. *NeuroImage*, 197:93–108.
- 1185 Nunez, M. D., Nunez, P. L., and Srinivasan, R. (2016). Electroencephalography (EEG):
1186 neurophysics, experimental methods, and signal processing. In Ombao, H., Linquist, M.,
1187 Thompson, W., and Aston, J., editors, *Handbook of Neuroimaging Data Analysis*, pages
1188 175–197. Chapman & Hall/CRC.
- 1189 Nunez, M. D., Srinivasan, R., and Vandekerckhove, J. (2015). Individual differences in
1190 attention influence perceptual decision making. *Frontiers in Psychology*, 8.
- 1191 Nunez, M. D., Vandekerckhove, J., and Srinivasan, R. (2017). How attention influences per-
1192 ceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters.
1193 *Journal of Mathematical Psychology*, 76:117–130.
- 1194 Nunez, P. L., Nunez, M. D., and Srinivasan, R. (2019b). Multi-Scale Neural Sources of
1195 EEG: Genuine, Equivalent, and Representative. A Tutorial Review. *Brain Topography*,
1196 32(2):193–214.

- 1197 Nunez, P. L. and Srinivasan, R. (2006). *Electric fields of the brain: the neurophysics of EEG*.
1198 Oxford University Press, Oxford ; New York, 2nd ed edition. OCLC: ocm58451867.
- 1199 Nunez, P. L., Wingeier, B. M., and Silberstein, R. B. (2001). Spatial-temporal structures
1200 of human alpha rhythms: Theory, microcurrent sources, multiscale measurements,
1201 and global binding of local networks. *Human Brain Mapping*, 13(3):125–164. eprint:
1202 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.1030>.
- 1203 O'Connell, R. G., Dockree, P. M., and Kelly, S. P. (2012). A supramodal accumulation-
1204 to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*,
1205 15(12):1729–1735.
- 1206 Ostwald, D., Porcaro, C., Mayhew, S. D., and Bagshaw, A. P. (2012). EEG-fMRI Based
1207 Information Theoretic Characterization of the Human Perceptual Decision System. *PLOS
1208 ONE*, 7(4):e33896.
- 1209 O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., and Kelly, S. P. (2018). Bridging Neural
1210 and Computational Viewpoints on Perceptual Decision-Making. *Trends in Neurosciences*,
1211 41(11):838–852.
- 1212 Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., and Turner, B. M.
1213 (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal
1214 of Mathematical Psychology*, 84:20–48.
- 1215 Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear
1216 analysis of EEG. *NeuroImage*, 28(2):326–341.
- 1217 Pfurtscheller, G., Stancák, A., and Neuper, C. (1996). Event-related synchronization (ERS)
1218 in the alpha band — an electrophysiological correlate of cortical idling: A review. *International
1219 Journal of Psychophysiology*, 24(1):39–46.
- 1220 Philiastides, M. G., Heekeren, H. R., and Sajda, P. (2014). Human Scalp Potentials Reflect a
1221 Mixture of Decision-Related Signals during Perceptual Choices. *Journal of Neuroscience*,
1222 34(50):16877–16889. Publisher: Society for Neuroscience Section: Articles.
- 1223 Philiastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural Representation of Task Diffi-
1224 culty and Decision Making during Perceptual Categorization: A Timing Diagram. *Journal
1225 of Neuroscience*, 26(35):8965–8975.
- 1226 Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using
1227 Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical
1228 Computing (DSC 2003)*, Vienna, Austria.
- 1229 Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., and Köthe, U. (2020). BayesFlow:
1230 Learning Complex Stochastic Models With Invertible Neural Networks. *IEEE Transac-
1231 tions on Neural Networks and Learning Systems*, pages 1–15. Conference Name: IEEE
1232 Transactions on Neural Networks and Learning Systems.

- 1233 Rangelov, D. and Mattingley, J. B. (2020). Evidence accumulation during perceptual
1234 decision-making is sensitive to the dynamics of attentional selection. *NeuroImage*, page
1235 117093.
- 1236 Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological review*,
1237 125(6):888.
- 1238 Ratcliff, R., Philiastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual
1239 decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the
1240 National Academy of Sciences*, 106(16):6539–6544.
- 1241 Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion Decision Model:
1242 Current Issues and History. *Trends in Cognitive Sciences*, 20(4):260–281.
- 1243 Regan, D. (1977). Steady-state evoked potentials. *JOSA*, 67(11):1475–1489. Publisher:
1244 Optical Society of America.
- 1245 Roitman, J. D. and Shadlen, M. N. (2002). Response of Neurons in the Lateral Intraparietal
1246 Area during a Combined Visual Discrimination Reaction Time Task. *Journal of Neuro-
1247 science*, 22(21):9475–9489.
- 1248 Ross, S. M. (2019). *Introduction to probability models*. Academic press.
- 1249 Rouder, J. N. and Morey, R. D. (2012). Default bayes factors for model selection in regression.
1250 *Multivariate Behavioral Research*, 47(6):877–903.
- 1251 Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python
1252 using PyMC3. *PeerJ Computer Science*, 2:e55.
- 1253 Schaworonkow, N. and Voytek, B. (2021). Enhancing oscillations in intracranial electro-
1254 physiological recordings with data-driven spatial filters. *PLOS Computational Biology*,
1255 17(8):e1009298. Publisher: Public Library of Science.
- 1256 Schmitt, M., Bürkner, P.-C., Köthe, U., and Radev, S. T. (2022). BayesFlow Can Reliably
1257 Detect Model Misspecification and Posterior Errors in Amortized Bayesian Inference.
- 1258 Schubert, A.-L., Nunez, M. D., Hagemann, D., and Vandekerckhove, J. (2019). Individual
1259 Differences in Cortical Processing Speed Predict Cognitive Abilities: a Model-Based
1260 Cognitive Neuroscience Account. *Computational Brain & Behavior*, 2(2):64–84.
- 1261 Shadlen, M. and Kiani, R. (2013). Decision Making as a Window on Cognition. *Neuron*,
1262 80(3):791–806.
- 1263 Srinivasan, R. (2004). Internal and external neural synchronization during conscious perception.
1264 *International Journal of Bifurcation and Chaos*, 14(02):825–842. Publisher: World
1265 Scientific Publishing Co.
- 1266 Srinivasan, R., Thorpe, S., and Nunez, P. L. (2013). Top-down influences on local networks:
1267 basic theory with experimental implications. *Frontiers in computational neuroscience*, 7:29.

- 1268 Swart, J. C., Frank, M. J., Määttä, J. I., Jensen, O., Cools, R., and den Ouden, H. E.
1269 (2018). Frontal network dynamics reflect neurocomputational mechanisms for reducing
1270 maladaptive biases in motivated action. *PLoS biology*, 16(10):e2005979.
- 1271 Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., and Van Maanen, L. (2017).
1272 Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical
1273 Psychology*, 76:65–79.
- 1274 Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., and Steyvers, M. (2016).
1275 Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *Neu-
1276 roImage*, 128:96–115.
- 1277 Twomey, D. M., Murphy, P. R., Kelly, S. P., and O’Connell, R. G. (2015). The classic
1278 P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*,
1279 42(1):1636–1643. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.12936>.
- 1280 van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derkx, K., Draws, T., Etz,
1281 A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, , Ly, A., Marsman,
1282 M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., and
1283 Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian
1284 analysis. *Psychonomic Bulletin & Review*, 28(3):813–826.
- 1285 van Ravenzwaaij, D., Brown, S. D., Marley, A., and Heathcote, A. (2020). Accumulat-
1286 ing advantages: A new conceptualization of rapid multiple choice. *Psychological review*,
1287 127(2):186.
- 1288 van Ravenzwaaij, D. and Etz, A. (2021). Simulation Studies as a Tool to Under-
1289 stand Bayes Factors. *Advances in Methods and Practices in Psychological Science*,
1290 4(1):2515245920972624. Publisher: SAGE Publications Inc.
- 1291 van Ravenzwaaij, D., Provost, A., and Brown, S. D. (2017). A confirmatory approach for
1292 integrating neural and behavioral data into a single model. *Journal of Mathematical Psy-
1293 chology*, 76:131–141.
- 1294 van Rooij, I. and Baggio, G. (2020). Theory Development Requires an Epistemologi-
1295 cal Sea Change. *Psychological Inquiry*, 31(4):321–325. Publisher: Routledge _eprint:
1296 <https://doi.org/10.1080/1047840X.2020.1853477>.
- 1297 van Vugt, M., Simen, P., Nystrom, L., Holmes, P., and Cohen, J. (2012). EEG Oscillations
1298 Reveal Neural Correlates of Evidence Accumulation. *Frontiers in Neuroscience*, 6:106.
- 1299 Vandekerckhove, J., White, C. N., Trueblood, J. S., Rouder, J. N., Matzke, D., Leite, F. P.,
1300 Etz, A., Donkin, C., Devezer, B., Criss, A. H., and Lee, M. D. (2019). Robust Diversity
1301 in Cognitive Science. *Computational Brain & Behavior*, 2(3):271–276.
- 1302 Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generaliza-
1303 tion of the savage-dickey density ratio. *Journal of the American Statistical Association*,
1304 90(430):614–618. 00422.

- 1305 Voss, A., Rothermund, K., and Voss, J. (2004). Interpreting the parameters of the diffusion
 1306 model: An empirical validation. *Memory & Cognition*, 32(7):1206–1220.
- 1307 Wabersich, D. and Vandekerckhove, J. (2013). *jags-wiener: a diffusion model plugin for*
 1308 *JAGS*.
- 1309 Wabersich, D. and Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom
 1310 distributions to JAGS (with a diffusion model example). *Behavior Research Methods*,
 1311 46:15–28.
- 1312 Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hy-
 1313 pothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive*
 1314 *psychology*, 60(3):158–189.
- 1315 Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., Broberg,
 1316 M., Wallace, A., DeLosAngeles, D., Lillie, P., Hardy, A., Fronsko, R., Pulbrook, A., and
 1317 Willoughby, J. O. (2007). Scalp electrical recording during paralysis: Quantitative evidence
 1318 that EEG frequencies above 20Hz are contaminated by EMG. *Clinical Neurophysiology*,
 1319 118(8):1877–1888.
- 1320 Wilson, R. C. and Collins, A. G. (2019). Ten simple rules for the computational modeling of
 1321 behavioral data. *eLife*, 8:e49547.
- 1322 Zhang, Q., van Vugt, M., Borst, J. P., and Anderson, J. R. (2018). Mapping working memory
 1323 retrieval in space and in time: A combined electroencephalography and electrocorticogra-
 1324 phy approach. *NeuroImage*, 174:472–484.

1325 8. Solutions

- 1326 1. Plotting of histograms are left to the reader. Plots of density approximations should
 1327 approximately reproduce the results in **Figure 2**. The code for this figure is given in
 1328 https://github.com/mdnunez/pyhddmjags/blob/master/simpleCPP_sim.py.
- 1329 2. A comparison model would be similar to Model 3 other than the fact that CPP slopes
 1330 c per trial i are not generated by drift-rates δ , a cognitive parameter that describes the
 mean rate of evidence accumulation.

$$(r_{ij}, x_{ij}) \sim DDM(\delta_j, \tau_j, \alpha_j, \beta_j, \eta_j) \quad (12)$$

$$c_{ij} \sim Normal(\phi_j, \sigma_j^2) \quad (13)$$

1329 Note that this comparison model has an additional parameter per participant, ϕ , com-
 1330 pared to Model 3. ϕ is just the mean CPP slope across trials.

- 1331 3. We could add an addition scaling parameter ψ in the second equation that can change
 1332 based on participant j :

$$c_{ij} \sim Normal(\psi_j \phi_j, \sigma_j^2) \quad (14)$$

- 1333 4. This model produces normally distributed response times, while Model 3 produces
 1332 response time distributions with right skews. Response time distributions with right
 1333 skews more realistically reflect real data.