Neuro-cognitive models of single-trial EEG measures describe latent effects of spatial attention during perceptual decision making

Ghaderi-Kangavari, A.; Rad, J.A.; Parand, K.; Nunez, M.D.

DOI
10.1016/j.jmp.2022.102725

Publication date
2022

Document Version
Final published version

Published in
Journal of Mathematical Psychology

Journal of Mathematical Psychology 111 (2022) 102725

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Neuro-cognitive models of single-trial EEG measures describe latent effects of spatial attention during perceptual decision making

Amin Ghaderi-Kangavari [a], Jamal Amani Rad [a,*], Kourosh Parand [a,b], Michael D. Nunez [c]

[a] Department of Cognitive Modeling, Institute for Cognitive and Brain Sciences, Shahid Beheshti University, Tehran, Iran
[b] Department of Data and Computer Sciences, Shahid Beheshti University, Tehran, Iran
[c] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 29 April 2022
Received in revised form 8 September 2022
Accepted 26 October 2022
Available online 16 November 2022

Keywords:
Perceptual decision-making
Spatial top-down attention
Single-trial EEG
Hierarchical Bayesian methods
Neuro-cognitive model
Drift–diffusion model

ABSTRACT

Visual perceptual decision-making involves multiple components including visual encoding, attention, accumulation of evidence, and motor execution. Recent research suggests that EEG signals can identify the time of encoding and the onset of evidence accumulation during perceptual decision-making. Although scientists show that spatial attention improves participant performance in decision making, little is known about how spatial attention influences the individual cognitive components that give rise to that improvement in performance. We found evidence in this work that both visual encoding time (VET) before evidence accumulation and other non-decision time processes after or during evidence accumulation are influenced by spatial top-down attention. Specifically, we used an open-source dataset in which participants were informed about the location of a target stimulus in the visual field on some trials during a face-car perceptual decision-making task. Fitting neural drift–diffusion models to response time, accuracy, and single-trial N200 latencies (∼ 125 to 225 ms post-stimulus) of EEG allowed us to separate the processes of visual encoding and the decision process from other non-decision time processes such as motor execution. These models were fitted in a single step in a hierarchical Bayesian framework. Quantitative model comparison to simulation-based theories reveals that spatial attention manipulates both VET and other non-decision time processes. We discuss why spatial attention may affect other non-evidence accumulation processes, such as motor execution time (MET), and why this may seem unexpected given the literature. We provide recommendations for future work to deal with this topic by a combination of neuro-cognitive models and model simulations at the single-trial level.

1. Introduction

In daily life, we frequently experience environments where we are obliged to make fast decisions via ambiguous and low-coherent sensory information (Bolam, Boyle, Ince, & Delis, 2022; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016). Perceptual tasks are useful for studying decision-making due to the precise control of the quantity and quality of sensory information as well as being able to manipulate latent underlying effects on response time and accuracy, such as visual attention (Gold & Shadlen, 2007). One of the most important factors that influence this decision process is top-down spatial attention. In everyday life, top-down prioritized factors can control visual attention, for example: knowledge, expectation, and current goals (Corbetta & Shulman, 2002; Philiastides, Diaz, & Gherman, 2017). In *spatial prioritization* tasks, used to explore the effects of top-down visual attention, a cue (e.g. arrow) in some experimental trials is often used to inform participants to attend covertly, without saccadic eye movements or head movements, to a location in the periphery of the visual field (Ostwald, Porcaro, Mayhew, & Bagshaw, 2012; Posner, 2016). These experiments translate directly to everyday experiences. For instance, when driving, expectations about the location of traffic lights help drivers more quickly execute the choice to stop or accelerate.

Visual perceptual decision making immediately after a visual stimulus is suspected to involve two classes of processes, decision process(es) containing an accumulation of evidence toward decision choices and/or urgency signals, and *non-decision time* (NDT) processes, thought to contain at least visual encoding time (VET) and motor execution time (MET) (Philiastides, Ratcliff, & Sajda, 2006; Ratcliff et al., 2016; Philiastides et al., 2017; Cisek, Puskas, & El-Murr, 2009; Evans, Hawkins, Boehm, Wagenmakers, & Brown, 2017; Nunez, Gosai, Vandekerckhove, & Srinivasan, 2019). Although non-decision time could contain

* Corresponding author.
E-mail addresses: amin.g.ghaderi@gmail.com (A. Ghaderi-Kangavari), j_amanirad@sbu.ac.ir (J.A. Rad), k_parand@sbu.ac.ir (K. Parand), m.d.nunez@uva.nl (M.D. Nunez).

https://doi.org/10.1016/j.jmp.2022.102725
0022-2496/

any non-evidence accumulation and non-urgency process. Top-down spatial cues are known to improve behavioral performance and manipulate neural mechanisms (Yeshurun & Carrasco, 1999; Ostwald et al., 2012; Sagar, Sengupta, & Sridharan, 2019). But while spatial prioritization is well studied, traditional proposed models could not separately identify different effects of spatial prioritization on VET, evidence accumulation, and other non-decision time components, such as MET simultaneously. In this study, we find evidence that spatial prioritization affects VET and other non-decision time components.

Two-alternative forced-choice tasks are described well by sequential sampling models. These models assume individuals accumulate adequate information until evidence is reached for one of two choices, typically conceptualized as hitting upper or lower boundaries (Ratcliff et al., 2016). Sequential sampling models often contain parameters with cognitive interpretations. These parameters can then be compared across experimental conditions to understand cognitive effects, as well as be compared directly to neural measures. For instance, researchers revealed that manipulating the task difficulty of the stimuli will specifically increase decision time by decreasing the drift rate, a parameter that tracks the average rate of evidence accumulation within a trial (Palmer, Huk, & Shadlen, 2005; Gomez, Ratcliff, & Childers, 2015; Ratcliff et al., 2016).

Event-Related Potentials (ERPs) are averages of EEG across experimental trials, time-locked to specific events such as the onset of a stimulus or the execution of a response, such as a button press. Even though early ERP responses from the brainstem after the onset of auditory stimuli may be presented within a few milliseconds, ERP responses from the primary visual cortex take approximately 40–60 ms (Luck, 2005). Visual evidence received by the retina must pass through the LGN to reach the primary visual cortex, then that information is preprocessed, decoded, and prepared before further cognitive use (Hall & Hall, 2020). Part of this process is target selection (Loughnane et al., 2016) while another component of this process is figure-ground segregation. The time course of figure-ground segregation is thought to depend upon visual elements such as distractors within visual stimuli and low coherence of stimuli (Lamme, Zipser, & Spekreijse, 2002; Nunez et al., 2019). We define the time between the onset of stimuli to the beginning of the accumulation process as *Visual Encoding Time (VET)*. VET is assumed to occur before the evidence can be accumulated during decision making, while some researchers show that evidence accumulation and motor planning could occur in parallel and not sequentially (Servant, White, Montagnini, & Burle, 2016; Dmochowski & Norcia, 2015; Verdonck, Loossens, & Philiastides, 2021).

VET is expected to finish between 150 ms and 225 ms after stimulus onset (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001; Nunez et al., 2019; Imani, Harati, Pourreza, & Goudarzi, 2021). For instance, when monkeys were trained to report the coherent direction of motion in a random dot motion task by a saccadic eye movement, groups of neurons in the lateral intraparietal cortex (LIP) were found to represent evidence accumulation for a saccadic choice (Roitman & Shadlen, 2002). The onset of this neural evidence accumulation typically begins around 200 ms after the presence of random dot motion stimuli (Roitman & Shadlen, 2002; Kiani & Shadlen, 2009; Shadlen & Kiani, 2013). Furthermore, participants in a go/no-go task required about 150 ms to process visual stimuli after stimulus onset (Thorpe et al., 1996).

Loughnane et al. (2016) identified two pairs of N200 ERP latencies, the peak time of negative deflections occurring in temporo-occipital electrodes around 200 ms after changes in bi-hemispheric visual stimuli. The authors showed that N200 latencies affect the onset of accumulated sensory evidence during a random dot motion task. Research has shown that N200 waveforms in posterior electrodes between 200 and 350 ms after stimulus onset reflect early visual information (Folstein & Van Petten, 2008). Recently, Nunez et al. (2019) revealed initial evidence that there is a 1 ms to 1 ms relationship between N200 components and VET. This evidence was derived from relationships between observed N200 latencies and estimated non-decision times related to figure-ground segregation before the accumulation process during a perceptual decision making task.

Over the last few years, researchers have begun to examine the underlying cognition and neural correlates of the decision making process simultaneously with neuro-cognitive modeling. These models allow consideration of underlying connections between cognitive model parameters and brain dynamics. Neuro-cognitive joint modeling is thought to be the most powerful technique for linking the electrophysiological dynamics of the brain across experimental trials to cognition and associated cognitive model parameters (Palmeri, Love, & Turner, 2017). This research has resulted in models that can join single-trial electroencephalography (EEG) measures and individuals' behavioral performance to make inferences about underlying states of the brain and behavior.

Three main model-fitting techniques have been used to relate unobserved brain and cognitive latent parameters, *Directed*, *Integrative* and *Covariance* approaches (Turner, Van Maanen, & Forstmann, 2015; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017; Turner, Forstmann, Steyvers, et al., 2019). In *Directed* approach, it is assumed that neural parameters constrain and predict cognitive latent parameters directly, and not vice versa (Nunez, Vandekerckhove, & Srinivasan, 2017; Wiecki, Sofer, & Frank, 2013; Frank, 2015). This approach is easily-understood and straightforward to implement, and the one we employed in this paper, in addition to direct simulations of the possible data generating models. This method allows us to "regress" one or more cognitive latent variables against single-trial EEG measures. We used this model-fitting procedure in a hierarchical Bayesian framework to estimate latent variables at individual and group levels. *Integrative* approaches can constrain and predict both cognitive and neural EEG parameters simultaneously (Palestro et al., 2018; Nunez, Vandekerckhove, & Srinivasan, 2022; Ghaderi-Kangavari, Rad, & Nunez, 2022). Both neural and behavioral data in this approach have been described with a set of shared parameters. *Covariance* approaches suggest that the cognitive and the neural parameters are constrained by a shared statistical distribution such as multivariate normal distribution (Turner et al., 2013, 2015; Palestro et al., 2018). In general, *Directed* approach assumes a one-way relationship between the brain and behavior, while both *Integrative* and *Covariance* approaches assume a two-way relationship between the brain and behavior. Properly accounting for single-trial EEG data in Integrative and Covariance models is difficult (Ghaderi-Kangavari et al., 2022). This is why we choose to use Directed models in this work in conjunction with simulations.

Using *Directed* single-trial EEG analysis, it has been shown that P200 measures after visual noise and N200 measures after visual stimuli (i.e. positive and negative deflections around 200 ms) could delineate single-trial visual attention effects on evidence accumulation and non-decision times (Nunez et al., 2017). Nunez et al. (2019, 2017, 2022) proposed neuro-cognitive hierarchical models of decision making to separate measures of the non-decision process by fitting models to both brain electrophysiology (EEG) and human response times and choices. These model parameters can then be related to visual attention as enforced by experimental paradigms.

In our previous work, we show that non-decision time is affected by spatial top-down attention in face-car perceptual decision-making task (Ghaderi-Kangavari, Rad, Parand, Ebrahimpour, & Nunez, 2021). In the current study, we hypothesized

that single-trial N200 peak latencies would reveal the effects of spatial attention on the non-decision process across experimental conditions and participants. We then sought to differentiate whether spatial prioritization influences VET or other non-VET non-decision times or both during perceptual decision making. Using a public dataset of a face-car perceptual decision-making task, we used singular value decomposition (SVD) to extract single-trial N200 latencies for all conditions (two levels of spatial prioritization and two levels of visual coherence) and then applied neurocognitive modeling to find associations between the Drift–Diffusion Model (DDM) parameters and the N200 latency on each experimental trial. We constructed hierarchical models to identify which components of NDT are the most influential to spatial top-down cues and then conducted model comparisons informed by a simulation study. We found evidence that spatial prioritization can affect other non-decision time processes in addition to VET while not affecting decision-making itself.

2. Methods

2.1. Data collection

We used a public dataset from an experiment to understand the interaction of perceptual decision making and spatial top-down attention (Georgie et al., 2018). In this work, seventeen participants (8 females, mean age was 25.9 years, range 20–33 years, 2 left-handed) from the University of Birmingham were recruited to perform a face-car perceptual decision-making task (see Fig. 1). Both behavioral data (response time and accuracy) and data from electroencephalograms (EEG) were recorded while participants performed the task.

The data collection was separated into two experimental sessions of approximately 10 min each. At the beginning of each trial, a one-way arrow cue (left or right) or two-way arrow cue was presented for one second, followed by a visual stimulus (face or car) that was shown for 200 ms. Participants are instructed to press a button based on whether they perceived a face or car stimulus. Participants used their index finger and middle finger of the right hand to respond. To avoid anticipatory responses by participants, an inter-stimulus interval (ISI) between 0 msec and 300 ms was used. During the task, all participants were instructed to maintain their gaze at the fixation point. This approach ensures the detection of spatial covert attention to stimuli rather than overt eye movements. The two different arrow types made up the spatial prioritization experimental conditions, with the "prioritized" level given by those trials with a one-way arrow and the "non-prioritized" level given by those trials with a two-way arrow. There was an additional experimental condition of phase coherence, reflecting stimulus informativeness and thus task difficulty, such that there was a manipulation of coherence of the visual stimuli (face or car stimulus), split into "low" and "high" coherence levels. Additional details about the definition and experimental manipulation of phase coherence are given in the Appendix and by Georgie et al. (2018). The two-level coherence manipulations and prioritized and non-prioritized cues were independent variables to manipulate the ambiguity of sensory information and spatial attention to a location of visual fields respectively.

One of the participants was excluded from the dataset because of both behavioral and EEG data corruption. All trials were randomly selected for each of four manipulations. For more information about experimental task design and procedure, readers can refer to the original work written on the data by Georgie et al. (2018).

Each participant performed the task for a total of 288 trials. This resulted in 72 trials for each unique combination of experimental manipulations: high coherence and prioritization, high coherence and non-prioritization, low coherence and prioritization, and low coherence and non-prioritization. For each manipulation, 36 trials for each face and car were randomly presented. For EEG data acquisition, 64-channels with 10–20 systems and two extra sensors relating to electrocardiogram (ECG) signals and correcting eye-blinking artifacts (EOG) were used. The ECG sensor was attached approximately 2 cm under the left collarbone, and the other sensor was placed under the left eye. Finally, if a response was faster than 150 ms, the trial was discarded resulting in 80 trials for all participants.

2.2. EEG preprocessing

EEG data is an amalgam of muscle and other biological artifacts, electrical noise, and true brain oscillations, therefore it needs to be cleaned to extract signal and related task fluctuations (Luck, 2005). We applied some oft-used preprocessing steps to the EEG data. However, some preprocessing stages related directly to N200 latency and SVD decomposition such as downsampling and band-pass filtering. These preprocessing steps were conducted using the MNE Python module (Gramfort et al., 2013).

The summary of preprocessing steps is as follows: (1) downsampling the raw data to 1024 samples per second from the original 5000 samples per second, (2) applying a Butterworth IIR band-pass filter from 1 to 10 Hz to match the filter parameters used by Nunez et al. (2019) to calculate single-trial N200s (note that a 1 Hz highpass could decrease the amplitude of slower and later potentials such as the P300), (3) re-referencing to the common average reference, splitting the EEG data into epochs -100 ms to 400 ms time-locked to the face/car onset and subtracting the average baseline potential, (4) running Independent Component Analysis (ICA) with the fastica algorithm (Hyvarinen, 1999) in order to inspect and remove artifactual components based on visual inspection (using the power spectrum and variance of the signal of the component), such as muscle, eye blinks, or eye movements, without removing the affected data portions, and (5) automatically removing Independent Components whose time courses matched EOG sensors or ECG sensors (such that these ICs reflected eye blinks and heart rhythm artifacts respectively), and finally (6) converting the data back into sensor space from Independent Component space. The EEG preprocessing code was implemented by the MNE package in Python (Gramfort et al., 2013).

2.3. Estimation of single-trial N200 waveforms

Calculating ERPs is easy to implement and results in large signal-to-noise ratios (Luck, 2014). Many researchers therefore study individual differences relating to variances of evoked potentials (e.g. see Schubert, Nunez, Hagemann, & Vandekerckhove, 2019; Kappenman, Farrens, Zhang, Stewart, & Luck, 2021). Traditional trial-averaged event-related potential (ERPs) have two main disadvantages: (1) missing potentially relevant information on each trial and (2) requiring a large number of participants to test the scientific hypothesis (Clayson, Carbine, Baldwin, & Larson, 2019; Boudewyn, Luck, Farrens, & Kappenman, 2018).

An alternative is to use information from EEG records on single trials (Bridwell et al., 2018). Although single-trial EEG analyses provide appropriate information across trials, these analyses are less robust to artifact and noise. This results in variability across trials that makes it difficult to draw inferences from these analyses. Also, analyses using single electrodes often result in potentials that are not observed above higher amplitude noise.

One method to estimate ERPs on single-trials is to use spatial filtering techniques. In general, spatial filtering techniques such

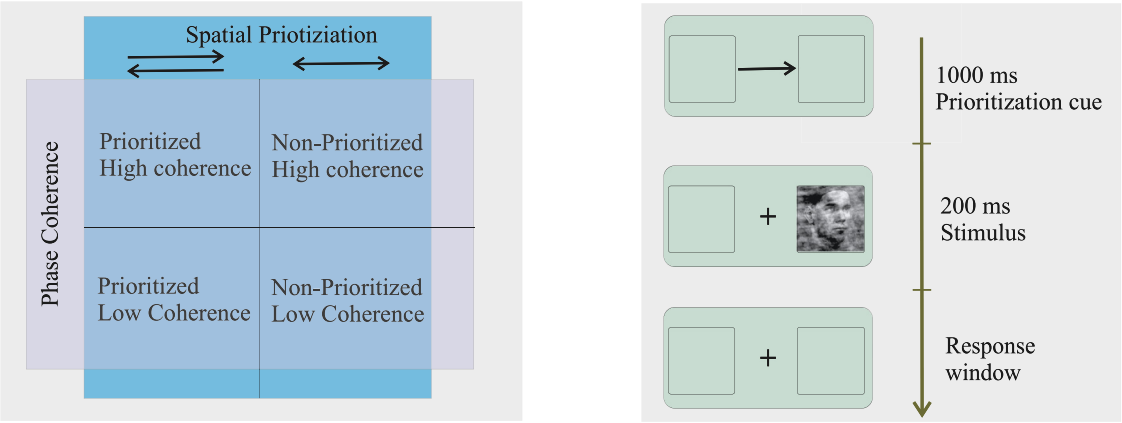


Fig. 1. Two levels were given to participants for coherence (high and low) and two levels for spatial attention (prioritized vs non-prioritized), randomly mixed across trials. For each trial, one-way arrows (informative cueing) or two-way arrows (uninformative cueing) were shown in the center of the screen for 1000 ms. Then, a picture of a car or face was shown on the left or right side of the screen for 200 ms. Participants then pressed a button to respond whether they perceived a car or a face. The study design was a 2 × 2 factorial design. Two levels were given to participants for coherence (high and low) and two levels for spatial attention (prioritized vs non-prioritized), randomly mixed across trials. For each trial, one-way arrows (informative cueing) or two-way arrows (uninformative cueing) were shown in the center of the screen for 1000 ms. Then, a picture of a car or face was shown on the left or right side of the screen for 200 ms. Participants then pressed a button to respond whether they perceived a car or a face. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as Independent Component Analysis (ICA) and Principal Component Analysis (PCA) can be applied to isolate a weighted feature associated with a cognitive process (e.g. spatial attention) (Cohen, 2014). Thus to mitigate problems with single-trial ERP analysis, we applied Singular Value Decomposition (SVD) to boost the signal-to-noise ratios for single-trial estimates (Nunez et al., 2017). This method uses a component of the trial-average ERP in all electrodes to find the optimal weights across electrodes to be used as a spatial filter to extract the N200s on single trials. This method produces a topological projection on the scalp for each individual relating to their waveform. Thus, one of the most important advantages of the SVD technique is that it does not require a predefined set of electrodes. Note that the SVD algorithm we used is non-stochastic and deterministic. Also note that PCA spatial filtering algorithms are closely related to Singular Value Decomposition (SVD) spatial filtering algorithms on both an algorithmic and empirical level, if not identical, depending upon whether the data were mean centered and the specific SVD/PCA algorithm (Shlens, 2014; Cohen, 2021).

By applying SVD to the trial-averaged data, we found N200 waveforms (time courses) and associated weight maps in either the first or second principal components (e.g. the two most influential components to the overall variance in the trial-averaged data). The input to the SVD algorithm was the matrix of ERPs consisting of 125 ms to 225 ms samples post-stimulus (see Fig. 2). Thus for each participant, the following formulas were constructed:

$$\mathcal{D}_{(T,C)} = \mathcal{U}_{(T,T)} \times \mathcal{S}_{(T,C)} \times \mathcal{V}^{*}_{(C,C)}, \tag{1}$$

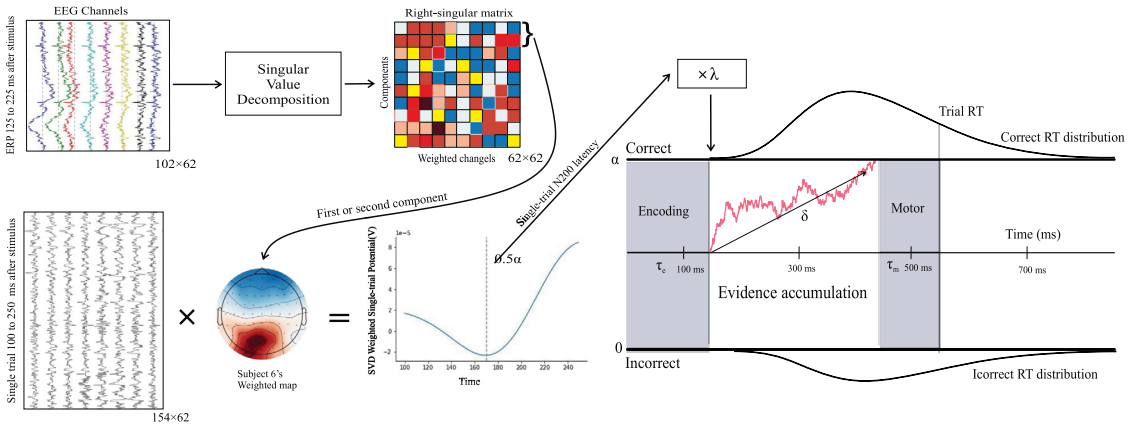
where $\mathcal{D}$ is a matrix of ERPs for each electrode, T is the number of time points between 125 to 225 ms after stimulus, which for our data is 102 points, C is the number of electrodes (62 electrodes). $\mathcal{U}$ contains the left-singular vectors, $\mathcal{S}$ is the matrix of singular values, and $\mathcal{V}$ contains the right singular vectors.

The first or second right singular vector v ($C \times 1$), corresponding to the first and second components respectively, was multiplied to the continuous data matrix E (from 125 to 225 ms after stimulus) for every single trial, then an $N \times 1$ vector $E \times v = e$ was obtained. This approach is different from the traditional ERP estimation method that computes potentials from single electrodes or pre-specified groups of electrodes, and not all electrodes. Our single-trial technique computes single-trial estimates using a weighted map of all electrodes. In order to compute minimum latencies of N200 waveforms, a few time measurement windows were considered by observing the distributions of N200 latencies and the N200 waveform (see Fig. 5) before using these values in further analysis. The following windows were explored by not ultimately used: 100–300 ms post-stimulus, 125–275 ms post-stimulus, 125–225 ms post-stimulus, 150–250 ms post-stimulus, and 175–274 ms post-stimulus. A 125–225 ms window post-stimulus of the ERP data was chosen to be submitted to SVD to maximally find N200 SVD components. Then a slightly larger window of 125–225 ms was used to compute the N200 minimum latency. Minimum values found at the boundaries of 125 or 225 ms of single-trial N200 peak-latencies were removed before neuro-cognitive model fitting since they were characteristic of a ramping potential on a trial rather than a clear N200 fluctuation. This resulted in removing a mean of 12.73% of trials (a maximum of 19% and a minimum 2%) across participants (see Fig. 5).

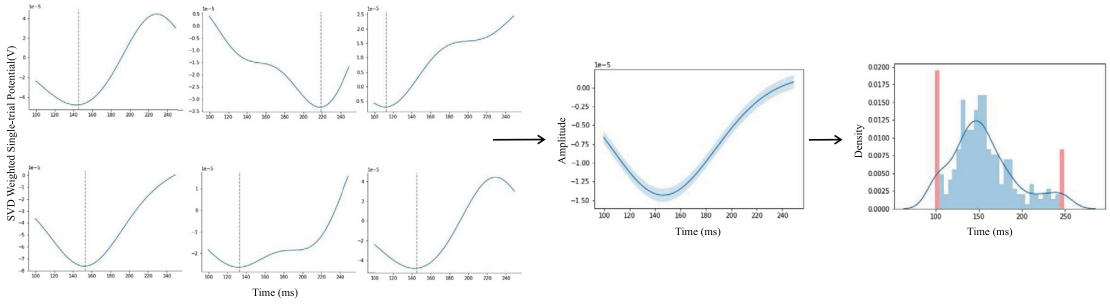
Note that we used an analysis with all electrodes to verify that N200 latencies were located in posterior electrodes. However, we could have considered only the posterior electrodes instead of the entire set of electrodes, matching the location of N200 waveforms in previous research (Folstein & Van Petten, 2008). In a post-hoc analysis suggested by a reviewer, we restricted SVD analysis to posterior electrodes to understand whether such an analysis could reduce the number of contaminant N200 latency estimates at the 125 or 225 ms boundaries. We chose 28 electrodes located at occipito–parietal electrodes and then applied the SVD analysis. This would have resulted in removing a slightly larger mean of 14.33% of trials (a maximum of 23% and a minimum 2%) across participants. A paired t-test showed that there was a significant, albeit small, difference between the original SVD method and post-hoc approach for reducing the boundary effect ($t = -2.56$, $p = 0.022$). We therefore decided to not rerun the subsequent analyses with the post-hoc method.

2.4. Integrated neuro-cognitive model fitting

Nunez et al. (2019) proposed a theory that N200s reflect noisy estimates of visual encoding time (VET). The current work uses the previous theory to identify the underlying mechanisms affected by top-down spatial attention. We show VET can be separated from other non-decision processes and evidence accumulation by fitting hierarchical neuro-cognitive Bayesian models (e.g. see Wiecki et al., 2013; Frank, 2015; Nunez et al., 2017)



(a) Graph of the method used to construct the single trial N200s and their effect on DDM's structure.



(b) Single participant ERP, N200 waveform, and distribution on all trials across all electrodes.

Fig. 2. A visual overview of how to derive EEG N200 latencies at the single-trial level and how these measures lead to constraining and predicting the onset of evidence accumulation parameter in the drift–diffusion model. Figures (a) show that we first extracted single-trial N200 latencies by applying SVD to find the most relevant and informative component called the weighted map to transfer EEG signals of all the electrodes into an empowered single waveform. The Collection of figures in part (a) represents *Directed* neuro-cognitive models which combine both behavioral and neural data simultaneously. Figures in part (b) depict some single-trial N200 latencies and distribution across all trials. In the middle figure, the shading band identifies standard error across trials. In the figure on the right, you can see the boundary effect caused by the noise, and we try to alleviate their effect through implantation. Boundary values of N200 distributions are specified by the red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in conjunction with simulation of the theory. This approach helps researchers to test hypotheses of trial-to-trial and individual differences with the help of neural activity and human behavior during forced-choice fast perceptual decision-making (see Fig. 2). We used an extended form of the neuro-cognitive drift–diffusion model by adding intrinsic trial-to-trial variability (η) for drift rate (Ratcliff et al., 2016) that was not related to trial-to-trial variability in N200 in any model, and four main parameters consist of boundary α indicating the speed-accuracy trade-off and the level of conservatism, drift rate δ indicating the mean slope of evidence accumulation and level of task difficulty, starting point $\beta = .5$ indicating bias of the accumulation process, non-accumulation time τ indicating both of encoding time and execution time. To test our hypothesis, we fit four different hierarchical neuro-cognitive models consisting of embedded linear connections between single-trial N200 latencies from EEG and the non-decision time parameter. We also fit a comparison model with linear connections between single-trial N200 latencies and all parameters. We found in simulation (see later text) that these models can separate underlying latent components of non-decision time and also answer the question of whether VET and/or other non-decision time components are shifted significantly by spatial prioritization.

Cognitive scientists typically focus on group-level differences due to manipulations and disorders. However, individual differences in spatial attention may unveil some latent aspects of the prioritization related to sub-components of non-decision time (Lee, 2011). Also, in an individual-level Bayesian framework, we would have no common hierarchical parameters among all participants. Hierarchical models often result in more accurate estimates of parameters due to "shrinkage" towards mean parameters (Gelman, Carlin, Stern, Dunson, Vehtari & Rubin, 2014).

We, therefore, fit *hierarchical* neuro-cognitive models to incorporate intrinsic differences in individuals' cognitive strategies and abilities when assessing whether spatial prioritization affects visual encoding time and other non-decision times. These hierarchical neuro-cognitive models also allowed us to identify individual differences in the relationship of single-trial N200s to non-decision times.

To fit each model to all participants' data simultaneously, we used the probabilistic programming language and Markov Chain Monte Carlo (MCMC) sampler Stan (Carpenter et al., 2017) within Python using the PyStan connector. For each model and bootstrap iteration (for the behavior-only models, see Appendix "Testing hypotheses on behavioral data" section), three MCMC chains were run to generate 10,000 samples from the joint posterior distribution of parameters. The initial 2000 samples were discarded as a burn-in phase to minimize the effect of initial values on posterior inference, and then a thinning parameter of 2 was used to result in 4000 posterior samples in each chain. The convergence of the Markov chains was assessed through visual inspection as well as by calculating the Gelman–Rubin statistic, R-hat, to ensure that the models had properly converged

Table 1
The results of model comparison for hierarchical neuro-cognitive models for imputed N200 latencies. Based on model selection criteria, *Model 4A* is a better model to describe the manipulation of spatial prioritization. Lower WAIC and -ELPD indicate better fits to data after accounting for model complexity. Note that -LPPD does not consider the model complexity. The first column means that in all models, we have considered that phase coherence (coher) manipulates the drift rate parameter (δ), as this manipulation is often reported, resulting in fewer models and saving time. The second column shows that spatial attention (spat) can manipulate residual parameters (r). The third column displays that spat can shift the effect parameter (λ). The fourth column shows that spat can affect the drift rate (δ). The fifth column shows that spat can manipulate the boundary decision (α) and finally the sixth column manifests that spat may manipulate all parameters.

	$\delta \sim coher$	$r \sim spat$	$\lambda \sim spat$	$\delta \sim spat$	$\alpha \sim spat$	$Full \sim spat$	WAIC	-LPPD	PWAIC	-ELPD
Model 1A	√						−19169	−9901	316	−9541
Model 1B	√				√		−19182	−9919	328	−9550
Model 1C	√			√			−19183	−9931	339	−9548
Model 2A	√	√					−19218	−9937	328	−9562
Model 2B	√	√			√		−19214	−9943	336	−9561
Model 2C	√	√		√			−19212	−9957	351	−9560
Model 3A	√		√				−19220	−9937	327	−9564
Model 3B	√		√		√		−19216	−9944	337	−9561
Model 3C	√		√	√			−19227	−9962	349	−9569
Model 4A	√	√	√				**−19231**	−9949	333	**−9571**
Model 4B	√	√	√		√		−19223	−9954	343	−9562
Model 4C	√	√	√	√			**−19231**	−9972	356	−9569
Model 4A Full	√					√	−19208	**−9984**	380	−9556

(Gelman, Carlin et al., 2014). Specifically, R-hat is a statistic that compares between-chain and within-chain variance. The collection of posterior samples from each chain was used to form one posterior sample of 12,000 samples for each parameter. All four models converged based on R-hat statistics that were less than 1.01 for all parameters in each model.

Model 1A assumes that the linear relationship between single-trial N200 latencies and single-trial non-decision times does not depend on spatial prioritization. This model is designed to test a hypothesis that spatial prioritization itself would only shift the single-trial N200 latencies and not any relationship to non-decision times nor any residual non-decision time component unrelated to N200 latencies. *Model 2A* assumes that spatial prioritization would only shift the relationship of single-trial N200 latencies to non-decision time and not the residual non-decision time. *Model 3A* assumes that spatial prioritization only affects the non-decision time unrelated to N200 latencies. Finally, *Model 4A* assumes that spatial prioritization may affect both non-decision time related to N200 latencies and non-decision time unrelated to N200 latencies. The check (tick) marks in Table 1 summarize the model differences. For all models, uninformative prior distributions were used, such that our model-fitting procedure was data-driven.

For *Model 1A* (see Fig. 3(a)): indices of i refer to experimental trials. Also, indices of j refer to participants, k refers to conditions, k_1 refers to phase coherence, and k_2 refers to spatial prioritization. None of the residual NDT (r) nor linear coefficient parameters (the influence of the single-trial N200 latencies on NDT; λ) were free to vary with spatial prioritization. Distributions of parameters for each participant i and condition k were as follows:

$$\left(\delta_{jk_1}|\mu_{(\delta)k_1}, \sigma_{(\delta)}\right) \sim \mathcal{N}\left(\mu_{(\delta)k_1}, \sigma^2_{(\delta)}\right), \mu_{(\delta)k_1} \sim \mathcal{N}(2, 4^2),$$
$$\sigma_{(\delta)} \sim \Gamma(1, 1). \tag{2a}$$
$$\left(r_j|\mu_{(r)}, \sigma_{(r)}\right) \sim \mathcal{N}\left(\mu_{(r)}, \sigma^2_{(r)}\right), \quad \mu_{(r)} \sim \mathcal{N}(.2, .4^2),$$
$$\sigma_{(r)} \sim \Gamma(.1, 1). \tag{2b}$$
$$\left(\lambda_j|\mu_{(\lambda)}, \sigma_{(\lambda)}\right) \sim \mathcal{N}\left(\mu_{(\lambda)}, \sigma^2_{(\lambda)}\right), \quad \mu_{(\lambda)} \sim \mathcal{N}(.5, 2^2),$$
$$\sigma_{(\lambda)} \sim \Gamma(.1, 1). \tag{2c}$$
$$\left(\alpha_j|\mu_{(\alpha)}, \sigma_{(\alpha)}\right) \sim \mathcal{N}\left(\mu_{(\alpha)}, \sigma^2_{(\alpha)}\right), \quad \mu_{(\alpha)} \sim \mathcal{N}(1, 2^2),$$
$$\sigma_{(\alpha)} \sim \Gamma(1, 1). \tag{2d}$$
$$\left(\eta_j|\mu_{(\eta)}, \sigma_{(\eta)}\right) \sim \mathcal{N}\left(\mu_{(\eta)}, \sigma^2_{(\eta)}\right) \in (0,), \mu_{(\eta)} \sim \mathcal{N}(1, 1^2),$$
$$\sigma_{(\eta)} \sim \Gamma(1, 1). \tag{2e}$$
$$\left(\mu_{(z)jk}|\mu_{(z)k}, \sigma_{(z)}\right) \sim \mathcal{N}\left(\mu_{(z)k}, \sigma^2_{(z)}\right) \in (0, .4), \mu_{(z)k}$$
$$\sim \mathcal{N}(.15, .1^2), \sigma_{(z)} \sim \Gamma(.1, 1). \tag{2f}$$
$$y_{ijk} \sim Wiener\left(\alpha_j, r_j + \lambda_j x_{ijk}, \beta, \delta_{jk_1}, \eta_j\right). \tag{2g}$$
$$x_{ijk} \sim \mathcal{N}\left(z_{(x)jk}, \sigma^2_{(x)}\right) \in (.101, .248), \quad \sigma_{(x)} \sim \Gamma(.1, 1). \tag{2h}$$

Normal ($\mathcal{N}$) prior and hyperprior distributions with average and variance parameters were used for non-variance variables. Gamma (Γ) prior and hyperprior distributions with shape and scale parameters were used for variance parameters. *Wiener* is the Wiener first passage time distribution for DDM (Wabersich & Vandekerckhove, 2014).

For *Model 2A* only residual (r) parameters were free to vary with spatial prioritization (k_2). Prior and likelihood distributions were the same as *Model 1A* besides the following lines that now varied with spatial prioritization (k_2):

$$\left(r_{jk_2}|\mu_{(r)k_2}, \sigma_{(r)k_2}\right) \sim \mathcal{N}\left(\mu_{(r)k_2}, \sigma^2_{(r)k_2}\right), \mu_{(r)k_2} \sim \mathcal{N}(0, .4^2)$$
$$\sigma_{(r)k_2} \sim \Gamma(.1, 1). \tag{3a}$$
$$y_{ijk} \sim Wiener\left(\alpha_j, r_{jk_2} + \lambda_j x_{ijk}, \beta, \delta_{jk_1}, \eta_j\right). \tag{3b}$$

For *Model 3A* only linear coefficient (λ) parameters were free to vary with spatial prioritization (k_2). Prior and likelihood distributions were the same as *Model 1A* besides the following lines that now varied with spatial prioritization (k_2):

$$\left(\lambda_{jk_2}|\mu_{(\lambda)k_2}, \sigma_{(\lambda)k_2}\right) \sim \mathcal{N}\left(\mu_{(\lambda)k_2}, \sigma^2_{(\lambda)k_2}\right), \mu_{(\lambda)k_2} \sim \mathcal{N}(0, 2^2),$$
$$\sigma_{(\lambda)k_2} \sim \Gamma(.1, 1), \tag{4a}$$
$$y_{ijk} \sim Wiener\left(\alpha_j, r_j + \lambda_{jk_2} x_{ijk}, \beta, \delta_{jk_1}, \eta_j\right). \tag{4b}$$

For *Model 4A* (see Fig. 3(b)) both the residual (r) and linear coefficient (λ) parameters were free to vary with spatial prioritization (k_2). Again, prior and likelihood distributions were the same as *Model 1A* besides the following lines that now varied with spatial prioritization (k_2):

$$\left(r_{jk_2}|\mu_{(r)k_2}, \sigma_{(r)k_2}\right) \sim \mathcal{N}\left(\mu_{(r)k_2}, \sigma^2_{(r)k_2}\right), \quad \mu_{(r)k_2} \sim \mathcal{N}(0, .4^2),$$
$$\sigma_{(r)k_2} \sim \Gamma(.1, 1), \tag{5a}$$
$$\left(\lambda_{jk_2}|\mu_{(\lambda)k_2}, \sigma_{(\lambda)k_2}\right) \sim \mathcal{N}\left(\mu_{(\lambda)k_2}, \sigma^2_{(\lambda)k_2}\right), \quad \mu_{(\lambda)k_2} \sim \mathcal{N}(0, 2^2),$$

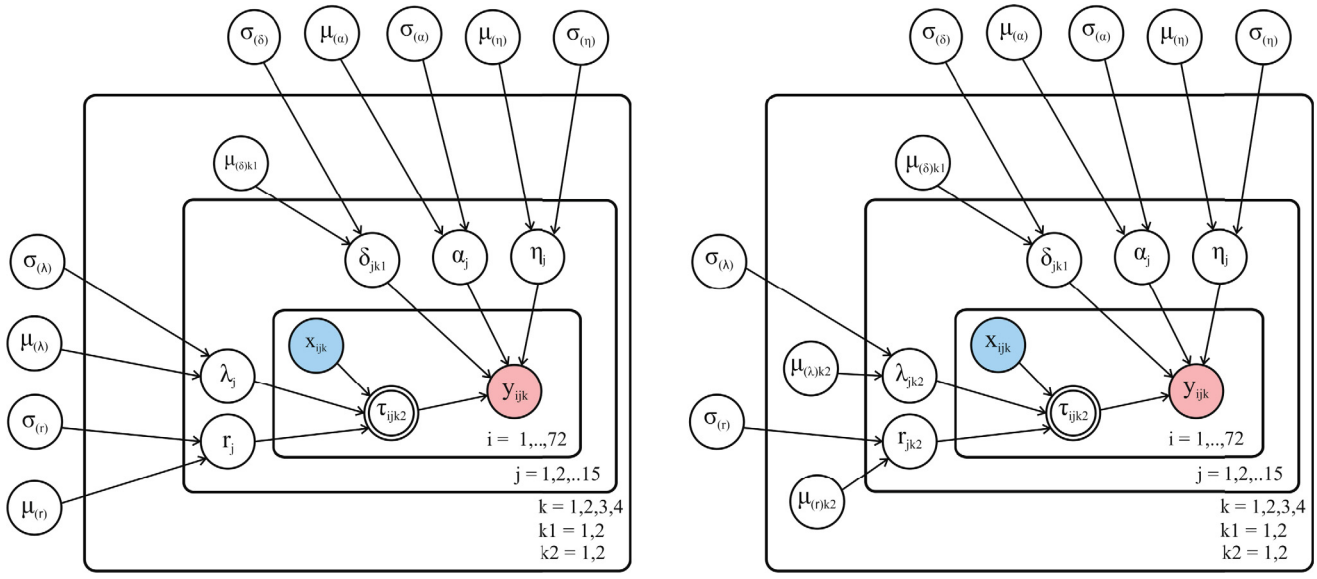


Fig. 3. Schematic diagrams of the hierarchical Bayesian models based on the convention of Lee and Wagenmakers (2014). Nodes show random variables in the model and arrows specify what variables affect other variables. The response time of y_{ijk} is an observed variable for the trial i level, the participant j level, and the condition k level. Indices of k_1 and k_2 display two-level coherence and two-level spatial prioritization. A double-bounded node in directed neurocognitive models refers to parameters that do not require estimation and are deterministic, meaning that their posterior distributions are directly calculated by the posterior distributions of other parameters.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\sigma_{(\lambda)k_2} \sim \Gamma(.1, 1), \tag{5b}$$

$$y_{ijk} \sim Wiener\left(\alpha_j, r_{jk_2} + \lambda_{jk_2}x_{ijk}, \beta, \delta_{jk_1}, \eta_j\right). \tag{5c}$$

We also fit a model in which all variables were free to vary by trial-to-trial N200 latencies. All parameters could change with the spatial attention condition k_2, but only the parameters related to drift rate could change with phase coherence. This model can be found in the Appendix "The full neuro-cognitive model, *Model 4A Full*" section and Fig. 11.

It was also possible that spatial cueing could affect the decision-threshold, such that top-down valid cues would lead to smaller decision-thresholds because participants would be less cautious in those trials. This alternative hypothesis was formally tested, using models similar to *Model 1A*, *2A*, *3A*, and *4A*, but with the decision-threshold parameter always changing by spatial prioritization (without any linear regressions of N200 latencies on decision-thresholds). To do so, we tested four other models labeled *Model 1B*, *Model 2B*, *Model 3B*, and *Model 4B* in which the α parameters in *Model 1A*, *Model 2A*, *Model 3A*, and *Model 4A* also varied by the spatial prioritization condition k_2. For instance, here is the specification of the likelihood for *Model 4B*:

$$y_{ijk} \sim Wiener\left(\alpha_{jk_2}, r_{jk_2} + \lambda_{jk_2}x_{ijk}, \beta, \delta_{jk_1}, \eta_j\right) \tag{6}$$

Moreover, we tested whether drift rate could be affected by spatial attention while non-decision time linearly covaried to trial-to-trial N200 latencies. For this purpose, we tested four models labeled *Model 1C*, *Model 2C*, *Model 3C*, and *Model 4C* in which the δ parameters in *Model 1A*, *Model 2A*, *Model 3A*, and *Model 4A* are free across all experimental conditions (by both phase coherence and spatial attention). The likelihood of such models is as follows:

$$y_{ijk} \sim Wiener\left(\alpha_j, r_{jk} + \lambda_{jk_2}x_{ijk}, \beta, \delta_{jk}, \eta_j\right) \tag{7}$$

where k belongs to $\{1, 2, 3, 4\}$ related two levels of spatial prioritization and two levels of coherence.

Each hierarchical neuro-cognitive model contained linear connections between single-trial N200 latencies and model parameters. Thus the relationship between DDM parameters and visual processing of the brain was evaluated in every trial. In this way, single-trial estimates of N200 latencies were assumed to be non-linearly related to choice-response times across all data points.

Note that these modeling assumptions were important to discover latent effects of spatial attention because manifest variables show little correlation. In the Appendix, we report a correlation Table 6 and scatter plot with a simple regression line in Fig. 10 of single-trial N200 latencies versus response times for each participant. Both provide inference about a lack of relationship between the manifest variables. We therefore felt that neuro-cognitive modeling was necessary to best estimate latent variables and understand the effects of spatial attention during perceptual decision making.

2.5. Imputing contaminant N200 latencies

We considered N200 latencies estimated at the boundaries to be contaminant N200 latencies (i.e. poor estimates of single-trial N200 latencies due to rising or decreasing potentials at the boundaries). We used the integrated neuro-cognitive model fitting procedure to impute these values, as they were likely considered Missing At Random (MAR) and not Missing Completely At Random (MCAR) (Barnard & Rubin, 1999). In every model, N200 latencies x_{ijk} were assumed to be generated from a normal distribution. These mathematical statements are given in the model statements above (e.g. the last distribution statement of *Model 1A*). The same mathematical statements were used to impute and replace contaminant N200 latencies.

Imputation can avoid biased parameter estimation when the data is MAR. However, we fit our models to both imputed and non-imputed data to check the robustness of our results in comparing models. That is, for the models with imputation we considered the generators of variables x_{ijk}. And for the models without imputation, we discarded the generators of variables x_{ijk}. The

findings (see Table 8 and Fig. 15 in the Appendix) confirms the robustness of the modeling results, such that both methods lead to the same conclusions.

2.6. Bootstrapping

Bootstrapping is a flexible statistical procedure that re-samples observed data to generate new simulated samples, and bootstrapping's simple implementation makes it attractive for use in psychological applications (Wagenmakers, Ratcliff, Gomez, & Iverson, 2004; Wright, London, & Field, 2011). Bootstrapping can have a variety of uses. For instance, researchers often apply bootstrapping to approximate confident intervals of parameters that are predicted by point estimation methods, such maximum likelihood method (Moore, McCabe, Duckworth, & Sclove, 2003). There are two different approaches to bootstrap. *Non-parametric Bootstrapping* uses only the observed data to re-sample and generates more data while *parametric bootstrapping* generates new samples from a parametric model which has first been fitted to the observed data. Both bootstrapping strategies have been previously used in maximum likelihood estimation (MLE) to construct difference distributions of goodness-of-fit for two different models and to differentiate between these models (Wagenmakers et al., 2004).

In this study, we used a *non-parametric bootstrapping* procedure to generate a new sample *of participants* with replacement 30 times from the observed data in hierarchical Bayesian models. We performed this analysis in order to improve model comparison inference and mitigate the influence of particular participants on model comparison. Note the power of our study to find the effects of spatial attention was based on the number of single trials data. For each iteration of model fitting, 30 participants were taken randomly with replacement from the 15 participants to estimate the posterior distribution of parameters.

In order to replicate and re-analyze the current behavioral dataset with a different approach from our previous work (Ghaderi-Kangavari et al., 2021), using bootstrapping, we built four models presented in Appendix part A and reported their results in Figs. 8 and 9. The results of model selection criteria such as WAIC, PWAIC, and ELPD show that non-decision time is a foremost parameter to explain spatial prioritization during perceptual decision making. We also performed bootstrapping for comparisons between *Model 1A*, *2A*, *3A* and *4A* in the current dataset. The results of these comparisons, in addition to model comparison for *Model4A Full* without using bootstrapping, are located in Table 3. For each model and bootstrap iteration, three Markov Chain Monte Carlo simulations (Gamerman & Lopes, 2006) were used to generate 10,000 samples from the joint posterior distribution of parameters, from which the initial 2000 samples were discarded as a burn-in phase to minimize the effect of initial values on the posterior inference and a thinning parameter of 2 resulting in 4000 posterior samples in each chain. The collection of posterior samples from each chain was used to form one posterior sample of 12,000 samples for each parameter. All four models converged based on R-hat were less than 1.01 for all parameters in each model. The original 16 participants were re-sampled with bootstrap to 30 participants.

2.7. Comparing models

Some model selection criteria were used to compare models and select the best models: the Widely Applicable Information Criterion (WAIC) such that lower WAIC are better, the log pointwise predictive density (lppd) is the log predictive accuracy of the fitted model such that higher is better, Expected Log Pointwise predictive Density for a new dataset (ELPD) such that higher is better, and the effective number of parameters (PWAIC) (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; Vehtari, Gelman, & Gabry, 2017; Fontanesi, Gluth, Spektor, & Rieskamp, 2019). According to the parsimony principle, if two models have the same explanation of the observed data, we should choose the simple model over the complex model, as evidenced by a lower PWAIC value (Vandekerckhove, Matzke, & Wagenmakers, 2015; Wagenmakers et al., 2004). The following equations have been used for calculating model selection criteria where S is the number of sample draws from the posterior distribution, $P(y_i|\theta_s)$ is the likelihood of individual sample s from 1 to S and each data point y_i from observed data $y_1, \ldots, y_N$, and $\widehat{Var}$ is variance (Tillman, Van Zandt, & Logan, 2020; Gelman, Hwang & Vehtari, 2014; Vehtari et al., 2017):

$$LPPD = \sum_{i=1}^{N} log\Big(\frac{1}{S}\sum_{s=1}^{S} P(y_i|\theta^s)\Big), \tag{8a}$$

$$PWAIC = \sum_{i=1}^{N} \widehat{Var}_{s=1}^{S}\Big(log(P(y_i|\theta^s))\Big), \tag{8b}$$

$$WAIC = -2\Big(LPPD - PWAIC\Big), \tag{8c}$$

$$ELPD = \sum_{i=1}^{N} \int p_t(\tilde{y}_i) log\Big(p(\tilde{y}_i|y)\Big) d\tilde{y}_i. \tag{8d}$$

2.8. Individual differences analysis

Hierarchical approaches assume that posteriors of participants' parameters come from a parent distribution such as normal distribution. This often leads to better parameter estimates due to shrinkage, even though the hierarchical distribution assumptions are not necessarily true (Gelman, Carlin et al., 2014). However, our neuro-cognitive hierarchical models had the disadvantage of containing potentially noisy EEG measures that differed across participants. In our models, it was difficult to incorporate these differences in EEG noise with hierarchical parameters. Therefore to ensure we picked the best model out of *Model 1A*, *Model 2A*, *Model 3A*, *Model 4A*, and *Model 4A Full*, we fit individualized (non-hierarchical) models to each individual and performed model comparison for each participant.

2.9. Simulation of three hypotheses and parameter recovery

To answer the question about whether spatial attention affects only visual encoding times or other non-decision times, or both, we considered four theories. These four theories were simulated directly and then the data was fit to each of the hierarchical neuro-cognitive models previously discussed. For each simulation, we simulated single-trial Visual Encoding Time (VET) and Motor Execution Time (MET). The specific single-trial VETs ($\tau_{(e)}$) and METs ($\tau_{(m)}$) were both assumed to come from a random draw from a normal distribution with a certain mean that could change based on the spatial prioritization condition and with a standard deviation of 100 ms (.1 s). We assumed that single-trial N200 latencies were drawn from a normal distribution with the mean of $.5\tau_{(e)i}$ and the variance of 50 ms. Note that N200 latencies are a purely visual ERP and have been shown to track the onset of evidence accumulation (Luck, 2005; Loughnane et al., 2016; Nunez et al., 2019). For this reason, we assumed that true VET only influenced single-trial N200 latencies positively, and that true MET (a reflection of all other possible non-decision times unrelated to visual processing) was unaffected by single-trial N200 latencies. In each of the three simulations, we varied which VET or MET would change with the two spatial prioritization (k_2)

levels, while the drift rate always changed by the two phase coherence levels (k_1).

The specific formulation of simulations was as follows:

$$\tau_{(e)i} \sim \mathcal{N}(\mu_{(\tau_e)k_2},\ .1^2), \tag{9a}$$

$$\tau_{(m)i} \sim \mathcal{N}(\mu_{(\tau_m)k_2},\ .1^2), \tag{9b}$$

$$x_i \sim \mathcal{N}(.5\tau_{(e)i},\ .05^2), \tag{9c}$$

$$y_i \sim \mathit{Wiener}(\alpha, \tau_{(e)i} + \tau_{(m)i}, \delta_{k_1}, \beta, \eta). \tag{9d}$$

In Simulation 1, we simulated N200 latencies named x_i and choice-response times named y_i with VET $\mu_{\tau_e} = .3$ (300 ms) in the prioritized level of the spatial prioritization condition and VET $\mu_{\tau_e} = .5$ (500 ms) in the non-prioritized level of the same condition. Also in Simulation 1, we simulated MET $\mu_{\tau_m} = .4$ (400 ms) in each level and thus MET (μ_{τ_m}) did not change by spatial prioritization. In Simulation 2, we simulated data with VET $\mu_{\tau_e} = .3$ in each spatial prioritization level but with MET changing across spatial prioritization levels. Such that in Simulation 2, MET $\mu_{\tau_m} = .4$ (400 ms) in the prioritized level and MET $\mu_{\tau_e} = .6$ (600 ms) in the non-prioritized level. In Simulation 3 both VET and MET changed, with VET $\mu_{\tau_e} = .3$ (300 ms) and MET $\mu_{\tau_m} = .4$ (400 ms) in the prioritized level and VET $\mu_{\tau_e} = .5$ (500 ms) and MET $\mu_{\tau_m} = .6$ (600 ms) in the non-prioritized level.

We also considered Simulation 4 in which the number of contaminated N200 measures changed across the prioritized and non-prioritized conditions. The following mixture formulations were used:

$$\tau_{(e)i} \sim \mathcal{N}(\mu_{(\tau_e)},\ .05^2), \tag{10a}$$

$$\tau_{(m)i} \sim \mathcal{N}(\mu_{(\tau_m)},\ .05^2), \tag{10b}$$

$$x_i \sim (1-\theta_{k_2}) \times \mathcal{N}(.5\tau_{(e)i},\ .05^2) +$$
$$\theta_{k_2}\mathcal{N}(.5\mu_{(\tau_e)},\ .5^2 \times .05^2 + .05^2) \tag{10c}$$

$$y_i \sim \mathit{Wiener}(\alpha, \tau_{(e)i} + \tau_{(m)i}, \delta_{k_1}, \beta, \eta). \tag{10d}$$

Note that one of the mixture distributions with probability $1 - \theta_{k_2}$ generates N200 latencies that vary with VET on single-trials i, while the other mixture with probability θ_{k_2} does not vary on single-trials. Therefore, we used this simulation to examine whether the quality of N200 latencies changed across spatial conditions due to increased artifact. Parameter θ changes by prioritized and non-prioritized conditions k_2. This simulation does not vary the variance of x_i, but it does manipulate the relationship between $\tau_{(e)i}$ and x_i. In fact, the simulation breaks the connection between the single-trial relationship of $\tau_{(e)i}$ and x_i on a percentage of θ trials. Note that the variance of the first and the second terms of the mixture model should be equalized which in this simulation both variances are $.5^2 \times .05^2 + .05^2$. The first mixture term assumes that x_i relates to VET values at the single-trial level, while the second term assumes that N200 latencies only reflect mean VET values and not on single trials (see **Model 4a** and **Model 4b** in Ghaderi-Kangavari et al., 2022). In Simulation 4, we simulated MET $\mu_{\tau_m} = .5$ (400 ms) in each level of spatial prioritization, VET $\mu_{\tau_e} = .3$ (300 ms) in each level, $\theta_{k_2} = .2$ in the prioritized level and $\theta_{k_2} = .4$ in the non-prioritized level.

For each of the four simulations we simulated the data with $\delta = 2.5$ (evidence units per second) in the high coherence level of the phase coherence condition and $\delta = 1.5$ in the low coherence condition. In each of the three simulations the other parameters were fixed across all levels of the two conditions: $\alpha = 1.5$ evidence units, $\beta = .5$ evidence units, $\eta = 0$ evidence units/s.

We then fit each of the four simulations to the first four hierarchical models. For each model fit three MCMC chains (Gamerman & Lopes, 2006) were used to generate 10,000 samples from the joint posterior distribution of parameters, from which the initial 4000 samples were discarded as a burn-in phase to minimize the effect of initial values on the posterior inference and a thinning parameter of 3 resulting in 6000 posterior samples. All models converged based on R-hat were less than 1.01 for all parameters in each model.

3. Results

3.1. Behavioral modeling results

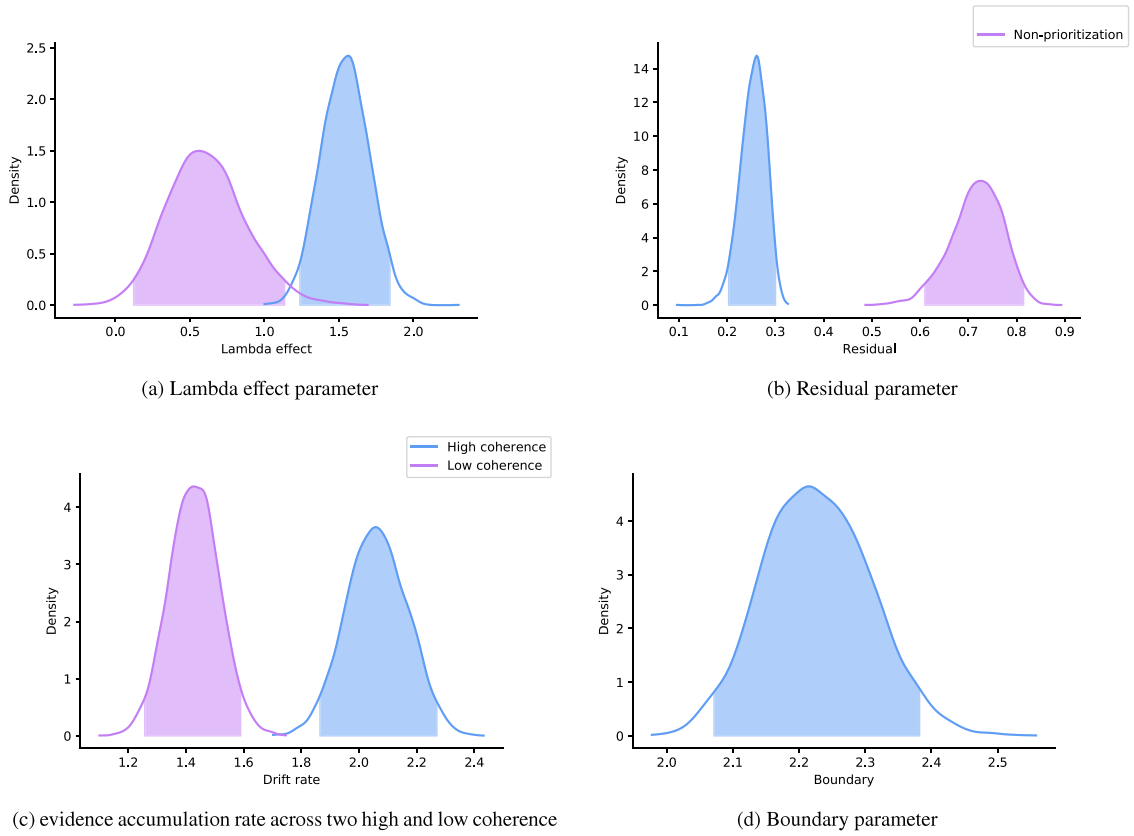
To understand the true effects of both independent variables, two-way repeated measure ANOVAs were used to reveal the true effects and interaction of the variables for both response time and accuracy. Fitting an ANOVA for mean response times across participants revealed significant main effects of phase coherence ($F(2, 14) = 31.72$, $p < 0.001$) and spatial cueing ($F(2, 14) = 14.29$, $p = 0.002$). These results imply that both the spatial cueing and phase coherence manipulations shifted response times for the face-car perceptual decision making task. In particular, the high-phase coherence condition had faster response times in comparison to low-phase coherence. Also, informative one-way arrows giving prioritized cues resulted in faster response times than when two-way arrows gave non-prioritized cues. Note that, for response time, the Cohen's d effect size of spatial manipulation (non-prioritization vs prioritization) and phase coherence (low coherence vs high coherence) are 1.5 and 1.01 respectively, and also for accuracy, the Cohen's d effect size of them are 2 and 0.47 respectively (Cohen, 2013). The interaction effect between phase-coherence and spatial cues was not significant ($F(2, 15) = 0.81$, $p = 0.38$). Note that because of these results we fit neuro-cognitive models with certain parameters that could change based on each experimental manipulation, reflecting the 2×2 task design, instead of allowing all parameters to change based on a four-condition design.

A two-way repeated measures ANOVA was also fit to mean accuracy across participants. The main effect of phase-coherence was significant ($F(2, 14) = 55.82$, $p < 0.001$), but the main effect of spatial cueing was not significant ($F(2, 13) = 3.0$, $p = 0.1$). Similarly the interactions of spatial cueing and phase coherence was not significant ($F(2, 13) = 0.05$, $p = 0.81$).

The results of cognitive modeling of behavior when fitting hierarchical DDMs across participants confirmed that non-decision time is the most influenced parameter by spatial prioritization (see Fig. 8 in Appendix part A). We used a bootstrap procedure with 30 draws. This was a direct replication of our previous work (Ghaderi-Kangavari et al., 2021). Specifically, to assess four cognitive models (see Appendix), we computed WAIC across bootstrap samples. For each iteration of the bootstrap sample, non-decision time was selected as the most likely parameter manipulated by spatial cueing. For each of the cognitive models, the means of WAIC, PWAIC, and ELPD across bootstrap samples are displayed in Appendix part A, Table 5. This result corresponds to the behavioral result because spatial cueing significantly changed mean response time and not mean accuracy. This is exactly what we would expect since this effect only influenced non-decision time and not evidence accumulation.

3.2. Parameter recovery in the simulated data

In order to understand how the model fits translates to the three simulated hypotheses, we checked both model comparison criteria and posteriors of parameter recovery for each of the three simulations. For Simulation 1, the fit of the data to *Model 1A* fits the data the best between *Model 1A*, *2A*, *3A*, and *4A* in terms of both WAIC and ELPD, see Table 2 and see Figs. 12 and 13. For Simulation 2, the fit of the data to *Model 2A* fits the data the best between *Model 1A*, *2A*, *3A*, and *4A*, see Table 2 and Fig. 14.



(a) Lambda effect parameter

(b) Residual parameter

(c) evidence accumulation rate across two high and low coherence

(d) Boundary parameter

Fig. 4. Parameter posteriors of *Model 4A* fit Simulation 3. The shaded area represents the 95% Bayesian credible interval (BCI) of posteriors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Model comparison criteria for three simulated participants.

Simulation	Criterion	*Model 1A*	*Model 2A*	*Model 3A*	*Model 4A*	*Model 4A Full*
1	WAIC	−298*	−294	−293	−286	**−304**
	ELPD	147*	145.0	145	140	**150**
2	WAIC	−967	**−1110***	−1087	−1105	−1063
	ELPD	482	**554***	543	551	530
3	WAIC	−419	−688	−575	−718*	**−726**
	ELPD	185	326	260	343*	**346**
4	WAIC	−1483	−1487*	−1484	−1484	**−1494**
	ELPD	741	742*	741	741	**745**

* Indicate the best model fit statistic between *Model* 1*A*, 2*A*, 3*A*, and 4*A*. Bold text indicates the best model fit statistic across all models (including *Model* 4*A Full*).

For Simulation 3, the fit of the data to *Model 4A* fit the data the best between *Model 1A*, *2A*, *3A*, and *4A*, see Table 2 and Fig. 4. For Simulation 4, the result of model fits to the simulated data shows that *Model 2A* is the best between *Model 1A*, *2A*, *3A*, and *4A*, see Table 2. Fig. 16 shows the parameter posteriors of *Model 4A* for Simulation 4.

Although the minimum WAIC and maximum ELPD are signs of the best model, *Model 4A Full* did not differentiate between the three hypotheses, with both Simulation 1 and Simulation 3 being best fit by *Model 4A Full*. Therefore true data results with *Model 4A Full* being the best fit model would not differentiate data between two hypotheses, namely that only VET or both VET and MET were affected by spatial prioritization. We did not find convergent evidence that *Model 4A Full* fit the data the best (see Table 1), and thus differentiated the hypotheses between the model fits of *Model 1A*, *2A*, 3 A and 4A.

3.3. Single-trial N200s were recovered for each participant with various levels of noise

We found single-trial N200s by using either the first or second component of the SVD (method described above). For each participant, we found component channel weights and component waveforms that best represented N200 latencies, see Fig. 5.

However, the results indicate that the amount of noise across trials in the single-trial estimation varied by participants. Specifically, the number of trials estimated at the boundaries of realistic N200 values varied among participants, suggesting that some single trials in those participants did not contain quality N200 latencies. In the neuro-cognitive models, we excluded these boundary effects by imputing N200 latencies for these trials while modeling choice-response times on these trials. Specifically, we used a Bayesian model as implemented in the Stan language, in which missing data were represented by new parameters.

We also analyzed the data of each specific participant in separate neuro-cognitive model fits without hierarchies across participants. In particular Participant 5 had the least boundary effects of single-trial N200 estimation and had clear posterior channel weights and a clear N200 waveform, therefore we considered the results of a neuro-cognitive model fits of only this participant, in addition to all participants (see Results below).

3.4. Spatial prioritization shifts both VET and MET

We wished to know which non-decision time sub-component was the most pivotal for manipulating spatial prioritization during face-car perceptual decision making. We used the results of four neuro-cognitive models to discover which VET or other non-decision time components were manipulated by a top-down spatial cue.

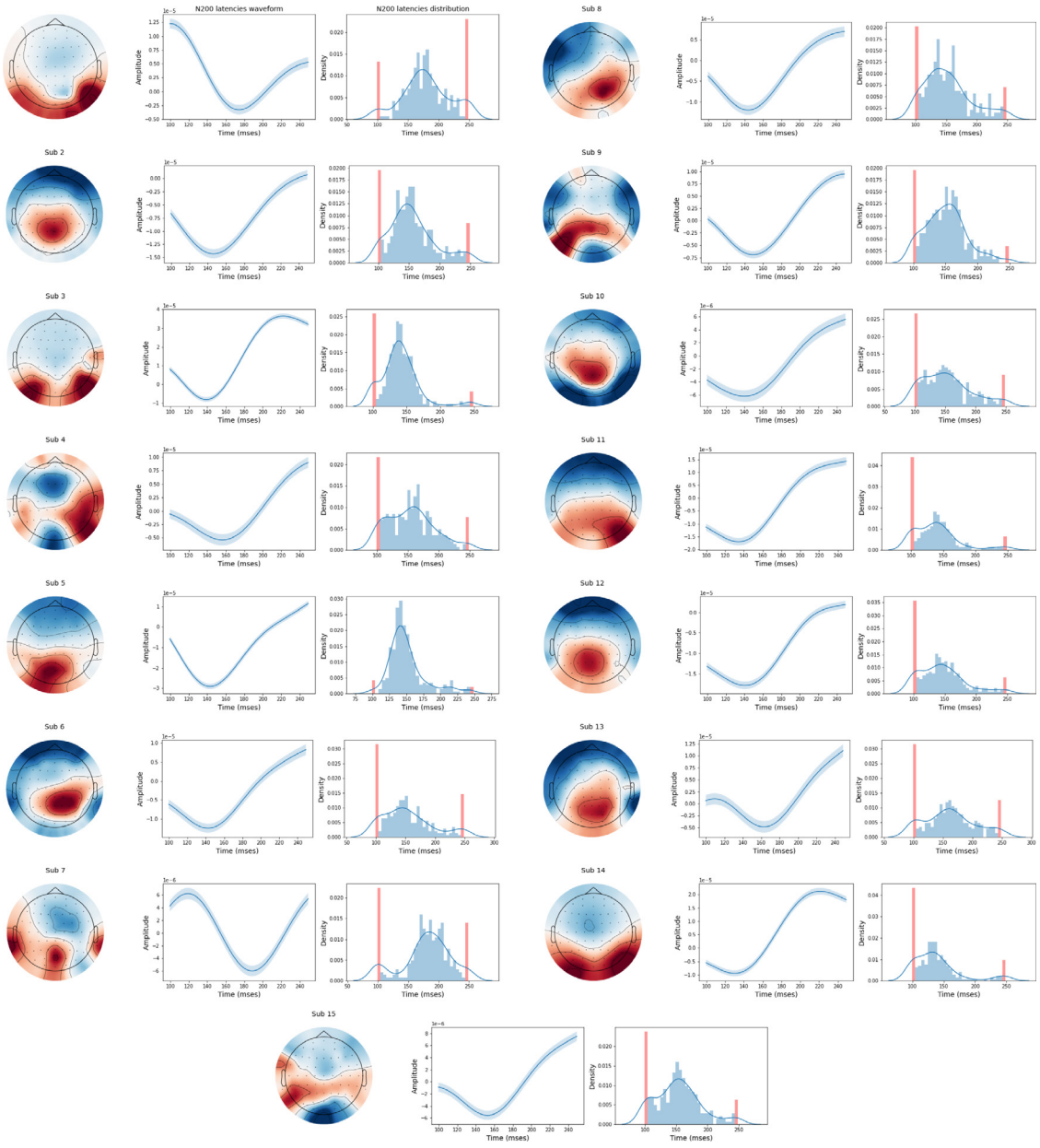


Fig. 5. Weighted map representations of the first or second SVD component of each participant. These scalp maps indicate the positive activation of occipito–parietal electrodes during the single-trial N200s peaks. The N200 waveform from the SVD method is displayed in the central column. and its pre-imputed density distribution on all trials across all electrodes for each participant. Boundary values of N200 distributions are specified by the red color in the histograms, these boundary values were removed and then imputed in each model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

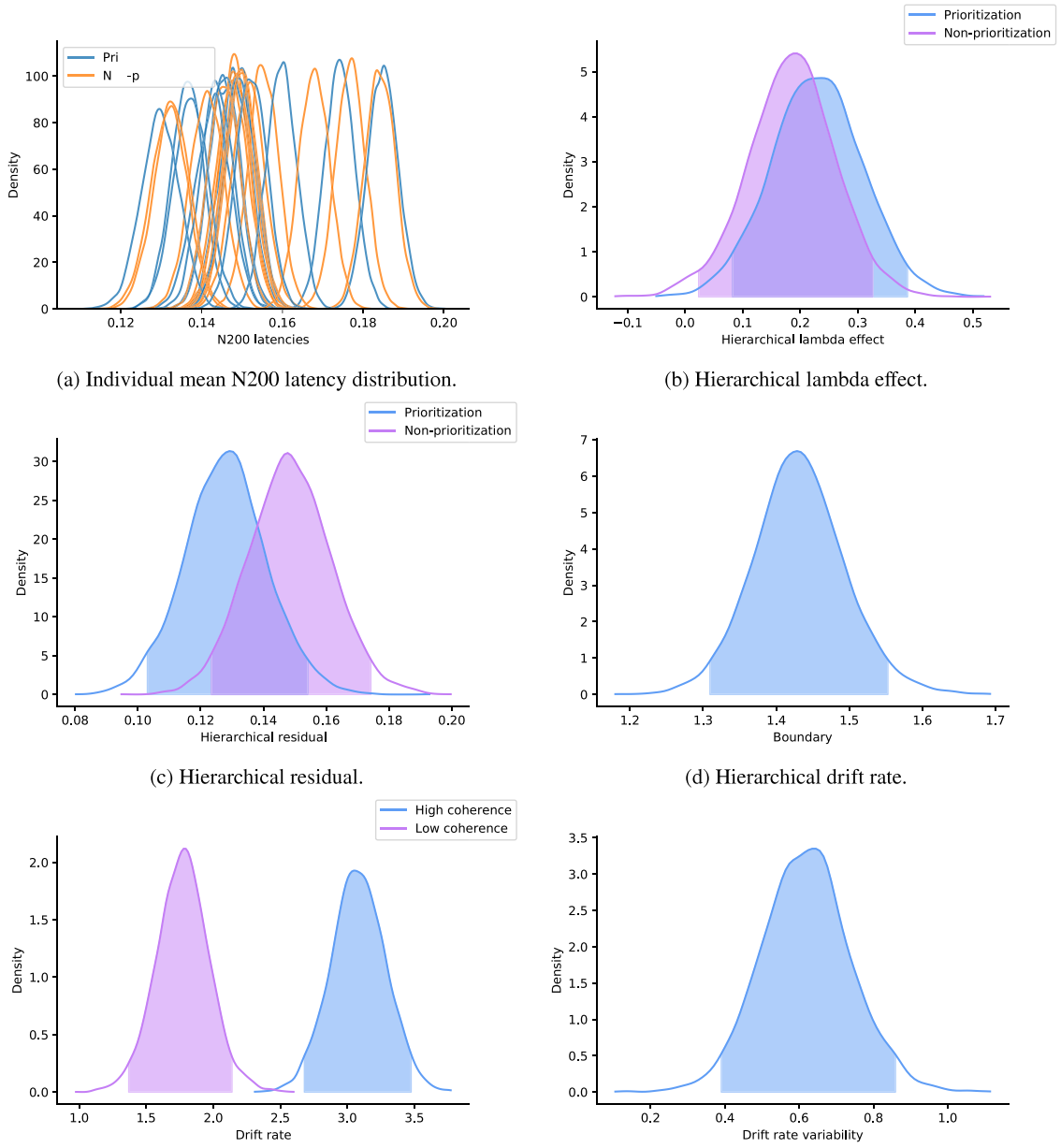
Based on model selection criteria in Tables 1 and 3, *Model 4A* is a better model to describe the manipulation of spatial prioritization. Lower WAIC and -ELPD indicate better fits to data after accounting for model complexity. Note that -LPPD does not consider the model complexity. For each of the bootstrap samples, WAIC and -ELPD of *Model 4A* are lower than the other models. For each of the neuro-cognitive models, the means of WAIC, PWAIC, and ELPD across bootstrap are given in Table 3.

Because *Model 4A* was selected as a better model of the data based on four model selection criteria, we report the results of the model's parameters and sub-component of its non-decision time, see Fig. 6. Of particular note is a difference in the posterior distributions of residuals r between the prioritization and non-prioritization levels.

We calculated the posterior probability of the difference in mean single-trial latencies between the non-prioritized $k_2 = 2$, and prioritized $k_2 = 1$ conditions, $z_{(x)j\{k_2=2\}} - z_{(x)j\{k_2=1\}}$, for each participant. To do so, we used posterior distributions of $z_{(x)jk}$ in the prioritized condition averaged across the phase coherence condition and then subtracted the posterior distributions of the non-prioritized condition, calculated using the same method. We used matched posterior samples to directly calculate the true uncertainty. We reported these probabilities for each participant in Table 7. We also performed t-tests of the differences. Both methods yielded similar inference. There was only some evidence for a difference in Participant 5 (95% confidence from the calculated posterior distribution).

Table 8 shows the results of fitting hierarchical neuro-cognitive models for non-imputed contaminant N200 latencies.

We also fit the neuro-cognitive models to each participant's data individually. We found that *Model 3A* fit 9 participants'



(a) Individual mean N200 latency distribution.

(b) Hierarchical lambda effect.

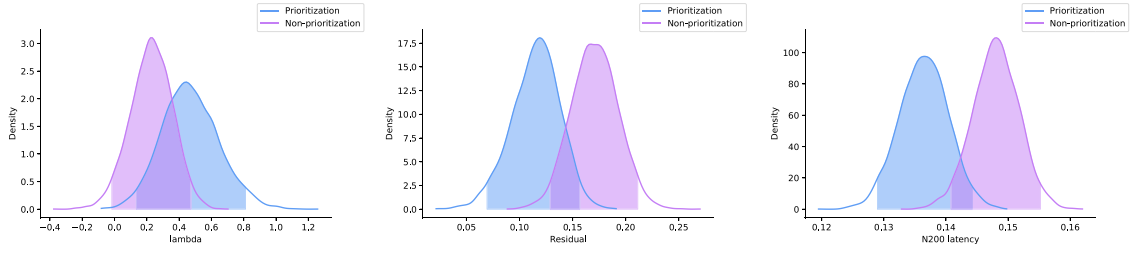
(c) Hierarchical residual.

(d) Hierarchical drift rate.

(e) Drift rate

(f) Hierarchical trial-to-trial drift rate variability

Fig. 6. Posterior distributions of Hierarchical Bayesian *Model 4A* for imputed N200 latencies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



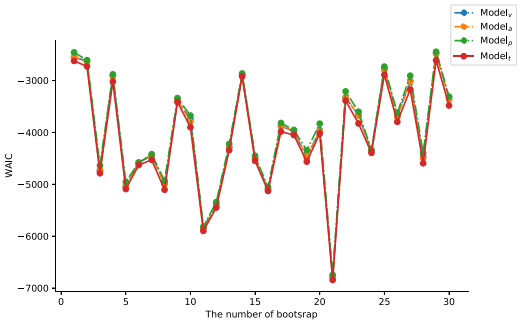
(a) Lambda effect.

(b) Residual parameter.

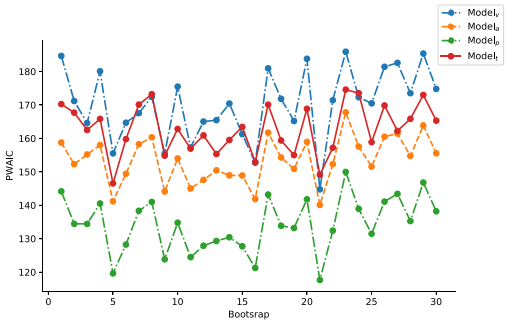
(c) Mean N200 latency.

Fig. 7. Posterior distributions of Hierarchical Bayesian *Model 4A* for imputed N200 latencies for participant 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

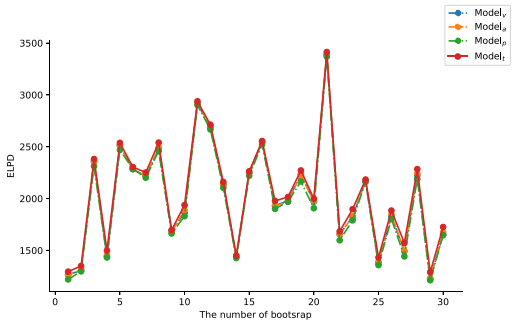
data the best. *Model 4A* fit 2 participants' data the best, and *Model 4A Full* fit 1 participant's data the best. However because the single-trial N200 results suggested that some participants had lower quality N200 estimation across trials (see Results in the section above), we also used the cleanest participant's single-trial N200 data to differentiate the three main hypotheses. Fitting the neuro-cognitive models to Participant 5 also resulted in *Model 4A* being the best model fit, see Table 4. Also, Fig. 7 shows posterior



(a) Watanabe-Akaike information criterion (WAIC)



(b) The effective number of parameters (PWAIC)



(c) Expected log pointwise predictive density (ELPD)

Fig. 8. The Four applicable information criteria of $M = 30$ bootstrap samples and four models. Each sample is the $n = 24$ double size with replacement from the 12 training samples. To choose the best model, the WAIC measure should be the lowest and ELPD should be the highest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Model comparisons across *Model* 1*A*, 2*A*, 3*A*, and 4*A* when fitting bootstrap samples. We implemented 30 bootstrap samples. For each bootstrap sample, we selected 30 participants at random with replacement. Each number shows the mean of criteria across all bootstrap samples. Lower WAICs indicate better fits to data after accounting for model complexity. The first column shows that in all models, we have assumed that phase coherence (coher) can manipulate drift rates (δ), as this manipulation is often reported, resulting in fewer models and saving time. The second column shows that spatial attention (spat) can manipulate residual parameters (r). The third column displays that spat can shift the effect parameter (λ). The fourth column shows that spat can manipulate the boundary decision (α) and finally the fifth column manifests that spat may manipulate all parameters. For other model selection criteria, larger values indicate better model fits data. Note that PWAIC is the number of effective parameters and thus is not a great measure of model comparison. Bold text indicates the best model fit statistic.

	$\delta \sim$ coher	$r \sim$ spat	$\lambda \sim$ spat	all	WAIC	PWAIC	-ELPD
Model 1*A*	√				−38463	632	−19137
Model 2*A*	√	√			−38566	655	−19185
Model 3*A*	√		√		−38566	660	−19183
Model 4A	√	√	√		**−38590**	669	**−19193**
Model 4A Full	√			√	−38567	726	−19182

parameters of the main N200 latencies, λ effect parameter and r parameter of *Model 4A* fit to Participant 5. This result reveals that spatial prioritization for a participant containing less artifact data creates clear manipulations of VET and MET.

4. Discussion

Accumulation models can split the cognitive time course of the perceptual decision making process into a decision time and a non-decision time. However, relying on fitting stochastic processes or formulations to behavioral data alone is not sufficient to decompose sub-components of non-decision time. Thus answering cognitive questions about the effects of experiential manipulations on non-decision processes during perceptual decision making requires more information than just participants' behavioral data. Scientists have previously used the electrophysiology activity to assist the quantitative models to be able to track latent psychological mechanisms better, including encoding and motor execution sub-components (Servant et al., 2016; Nunez et al., 2019; Weindel, Anders, Alario, Burle, et al., 2021). Neuro-cognitive models of drift–diffusion and single-trial EEG measures are a candidate to separate independent components of the decision process and to test hypotheses or differentiate theories (Nunez et al., 2022). In this study, we used assumed single-trial N200-latencies from occipito–parietal electrodes that dynamically track visual encoding time to understand the effects of spatial attention on visual encoding and other non-decision time components.

4.1. Brain networks of spatial attention

We found evidence that Visual Encoding Time (VET), as encoded by relationships of single-trial N200 latencies to response time distributions, is affected by spatial attention. Specifically, these single-trial N200 latencies were found in posterior electrodes in EEG, time-locked to the onset of visual stimuli. These results correspond to a large body of previous work to find spatial attention networks. A dorsal fronto-parietal network has been reported to play an effective role in top-down spatial attention (Chica, Bartolomeo, & Lupiáñez, 2013). This network usually extends to the occipital cortex (specifically, V1 and extrastriate areas), which likely reflects a top-down cue to modulate sensory representations (Moran & Desimone, 1985; Brefczynski & DeYoe, 1999). Bilateral BOLD signals in the primary visual cortex have been shown to be modulated with spatial cueing such that responses were greater when the subject attended to the stimuli in the contralateral hemifield (Gandhi, Heeger, & Boynton, 1999; Kastner, De Weerd, Desimone, & Ungerleider, 1998). Findings

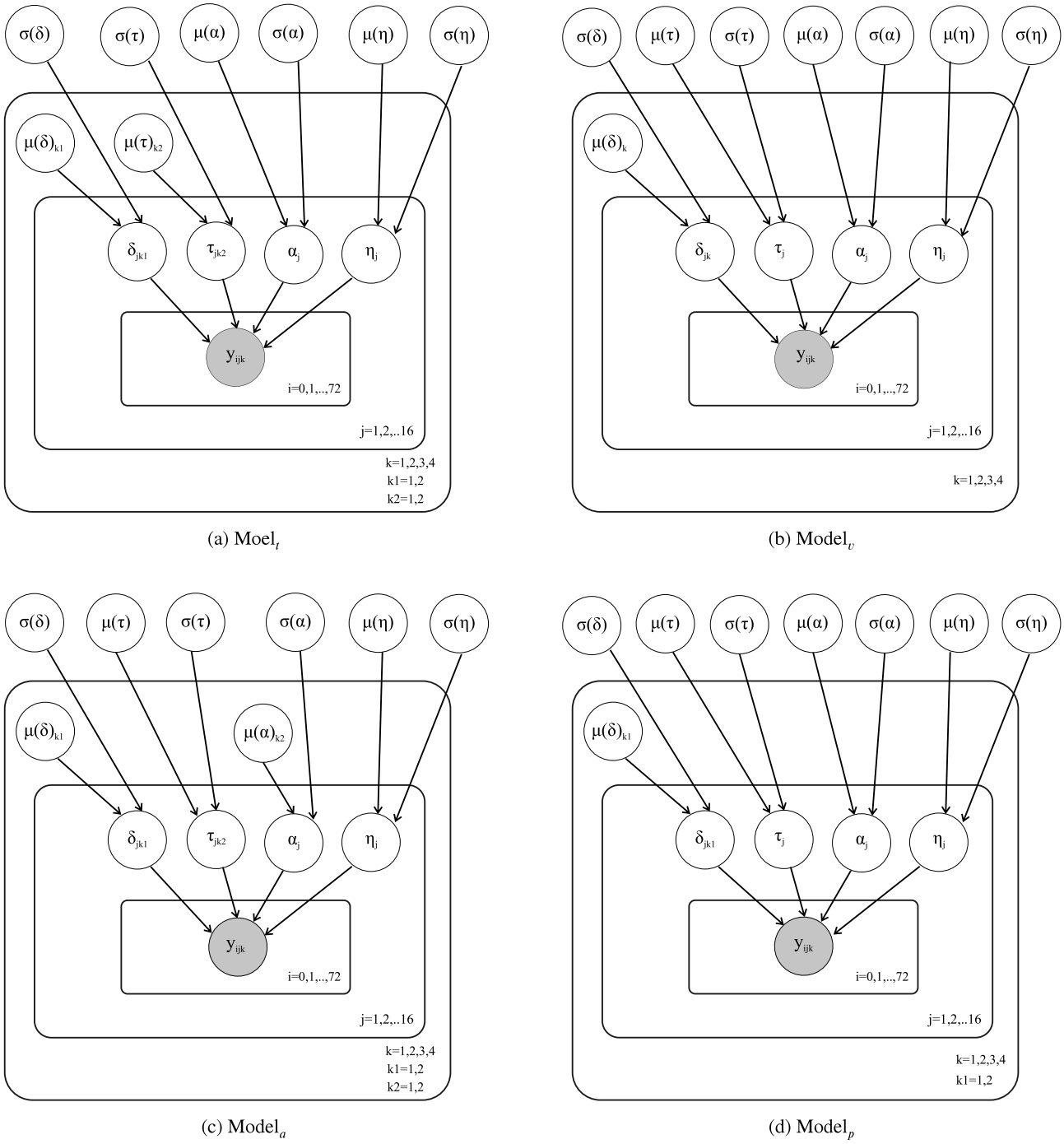


Fig. 9. Schematic diagrams of the four hierarchical Bayesian models based on the convention of Lee and Wagenmakers (2014). Nodes show random variables in the model and arrows specify what variables affect other variables. The response time of y_{ijk} is an observed variable for the trial i level, the participant j level, and the condition k level. Indices of k_1 and k_2 display two-level coherence and two-level spatial prioritization. In Model$_t$ amounts of participant-level parameters such as evidence accumulation rate δ_{jk_1}, non-decision time τ_{jk_2}, boundary separation α_j and trial-to-trial drift rate variability η_j, conditional hierarchical-level parameters $\mu(\delta)_{k_2}$ and $\mu(\tau)_{k_1}$, and non-conditional hierarchical-level parameters such as $\sigma(\delta)$, $\sigma(\tau)$, $\mu(\alpha)$, $\sigma(\alpha)$, $\mu(\eta)$ and $\sigma(\eta)$ are estimated by MCMC sampling. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

show that successive locations of visual targets lead to sustained activation of the intraparietal sulcus (IPS) and momentary activation of the occipital cortex (Corbetta & Shulman, 2002; Chica et al., 2013). The posterior parietal cortex consisting of the superior parietal lobule (SPL) and IPS is also a key component of an endogenous orienting network (Molenberghs, Mesulam, Peeters, & Vandenberghe, 2007). Finally, pre-stimulus alpha (8–14 Hz) oscillatory EEG activity in the parieto-occipital site is modulated by the endogenous spatial cueing paradigm (Worden, Foxe, Wang, & Simpson, 2000; Thut, Nietzel, Brandt, & Pascual-Leone, 2006; Sauseng et al., 2005).

However, we also found evidence that spatial attention affects additional non-decision times apart from Visual Encoding Time (VET). One possible reason is that Motor Execution Time (MET) is affected by spatial attention in this data. Before the study, we felt that the effect of spatial attention on MET was improbable. However, our results can be placed in the context of previous work. For instance, top-down signals for attending such as lateral

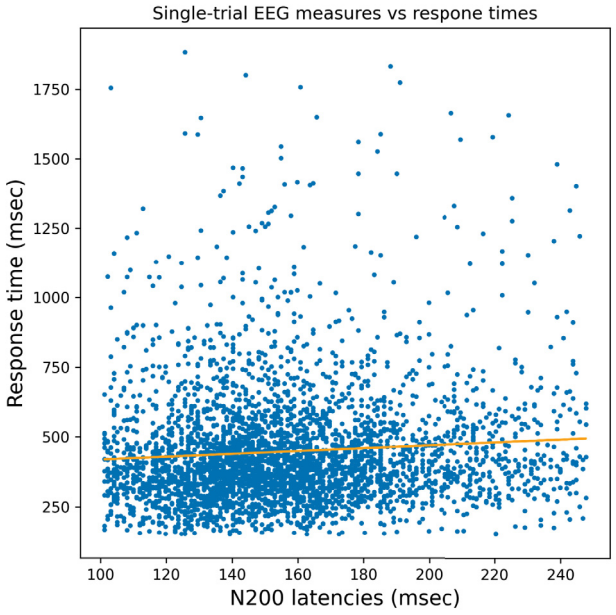


Fig. 10. A scatter plot of N200 latencies versus response time across all participants. Note that outlier single-trial N200 latencies on the window boundaries were removed before calculating the correlation coefficients. The best fit linear regression line between N200 latencies and response times ($F = 21.61$, and R-adjusted $= 0.006$ and p-value < 0.001) is shown in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

intraparietal area (LIP) have some effectors on response or action selection (Corbetta & Shulman, 2002). Attention-related modulations seem to happen primarily in response-related processes, and the collection of evidence shows that the primate premotor cortex including the dorsal premotor cortex (PMd) and supplementary motor area (SMA) plays a role in spatial attention (Simon et al., 2002; Klein, 1994; Macaluso & Driver, 2001). Finally, the premotor theory of attention derives from neurophysiological studies indicating that showing that spatial attention programs motor planning (Sheliga, Riggio, & Rizzolatti, 1995; Hunt, Reuther, Hilchey, & Klein, 2019).

4.2. Top-down spatial cue prioritization does not affect evidence accumulation and thresholds

Researchers have previously found evidence that fixation point (gaze) leads to an amplifying effect on the attended option (Krajbich, 2019). This phenomenon can be modeled with a sequential sampling theory of perceptual decision making, the attentional Drift Diffusion Model (aDDM) (Krajbich, Lu, Camerer, & Rangel, 2012). The assumption of the attentional Drift Diffusion Model (aDDM) is that attention can influence the rate of the evidence being gathered from the options during decision making, thus attention may shift the drift-rate (*delta*) parameter (Krajbich et al., 2012; Gwinn, Leber, & Krajbich, 2019). Also, the aDDM assumes that selective gaze will result in relevant information being accumulated while irrelevant information is ignored. This simple assumption is not necessarily correct, because subjects may gaze at a low-value option while *attending* to the high-value option, resulting in gathering more evidence in favor of a high-value option.

However, covert top-down spatial cue prioritization in the current experiment has a different mechanism to amplify performance. In this spatial attention manipulation, participants are only informed about the location of stimuli presence, and therefore top-down attention should not be related to dynamically shifting attention or eye movement between options during options in decision task. There is only one stimulus to which participants should attend, the face or the car stimulus. Thus after seeing the prioritization cue, participants have to make a decision between a face or a car in a certain level of phase coherence. It is therefore understandable that the most important latent parameter that is affected by spatial cueing cannot be drift rate in the current task (see top left of Fig. 11 in the Appendix).

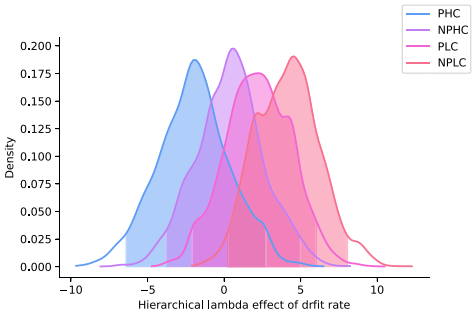
We proposed some nested models (scenarios) to test which parameters and subcomponents are influenced by spatial attention. Finally, after comparing hierarchical models we revealed that *Model 4A* best fit the data, which in the simulation was the best model when both VET and other non-decision time components are shifted by spatial attention. Hierarchical effects of single-trial N200 latencies were positive indicating relationships to VET, see Fig. 11(g). The remaining non-decision time is assumed to be an approximate measure of motor execution time (MET) or a mixture of other non-decision time processes. We found evidence that this non-decision time was also manipulated by spatial cueing, such that the prioritization level resulted in faster non-VET non-decision times than the non-prioritization level. Also, the model fit related to Simulation 4 reveals that N200 latencies across spatial prioritization cannot arise from artifacts and shows some more evidence that VET and MET are affected by spatial prioritization and not the N200 quality.

Note that the findings obtained by fitting neuro-cognitive models may improve the findings obtained from behavioral models. The outcome of behavioral models can restrict the variety

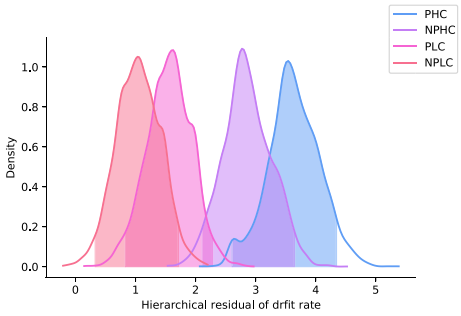
Table 4
Model comparison for individual-level Bayesian analysis to compare five nested models.

Participant	Criterion	*Model 1A*	*Model 2A*	*Model 3A*	*Model 4A*	*Model 4A Full*
1	WAIC	−1488	−1522	**−1525**	−1522	−1519
	ELPD	742	759	**761**	759	757
2	WAIC	−1046	−1047	**−1057**	−1048	−1050
	ELPD	522	523	**528**	523	523
3	WAIC	−1209	−1198	**−1221**	−1208	−1205
	ELPD	604	599	**610**	603	602
4	WAIC	−1406	−1409	**−1423**	**−1423**	−1414
	ELPD	703	704	**711**	**711**	707
5*	WAIC	−1362	−1361	−1362	**−1369**	−1358
	ELPD	680	680	680	**683**	679
6	WAIC	−1030	−1028	**−1056**	−1038	−1036
	ELPD	513	512	**525**	516	515
7	WAIC	−1520	**−1531**	−1530	−1530	−1527
	ELPD	756	**763**	761	762	760
8	WAIC	−1301	−1304	**−1345**	−1310	−1315
	ELPD	650	651	**672**	654	657
9	WAIC	**−1589**	**−1589**	**−1589**	−1587	−1585
	ELPD	**793**	**793**	**793**	792	791
10	WAIC	−540	−537	**−611**	−556	−570
	ELPD	269	267	**304**	277	283
11	WAIC	−1579	−1579	**−1593**	−1580	−1578
	ELPD	783	783	**790**	783	781
12	WAIC	−1237	−1241	**−1256**	−1240	−1239
	ELPD	616	618	**626**	617	617
13	WAIC	−750	−751	−753	**−756**	−751
	ELPD	375	375	376	**378**	375
14	WAIC	−1070	−1063	**−1235**	−1101	−1132
	ELPD	530	526	**612**	545	560
15	WAIC	−1152	−1174	−1174	−1171	**−1178**
	ELPD	574	585	585	584	**586**

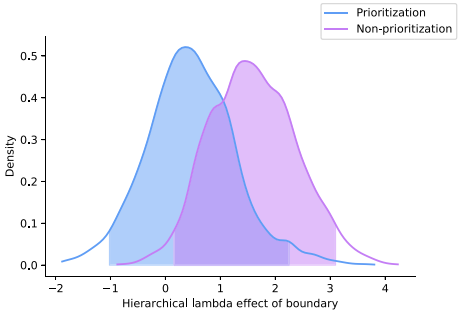
* Indicates the participant with the most trustworthy single-trial N200 results. Bold text indicates the best model fit statistic.



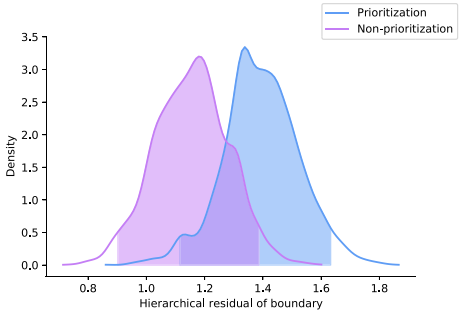
(a) Hierarchical lambda effect distribution of drift rate across conditions.



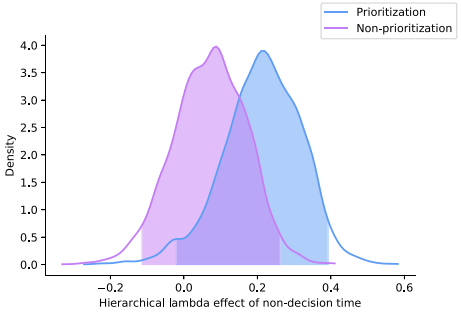
(b) Hierarchical residual distribution of drift rate across conditions.



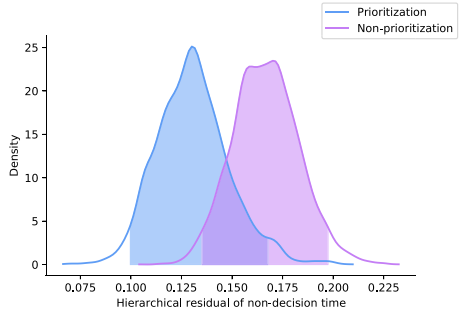
(c) Hierarchical lambda effect distribution of boundary across spatial attention.



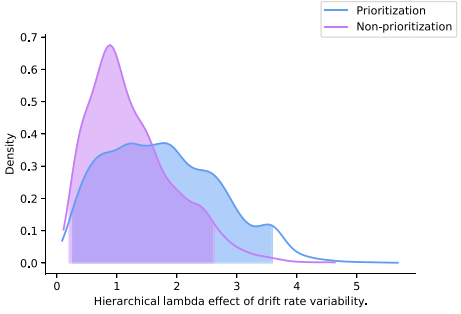
(d) Hierarchical residual distribution of boundary across spatial attention.



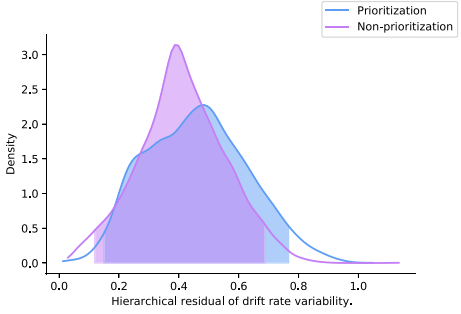
(e) Hierarchical lambda effect distribution of non-decision time across spatial attention



(f) Hierarchical residual distribution of non-decision time across spatial attention.



(g) Hierarchical lambda effect distribution of trial-to-trial variability across spatial attention.

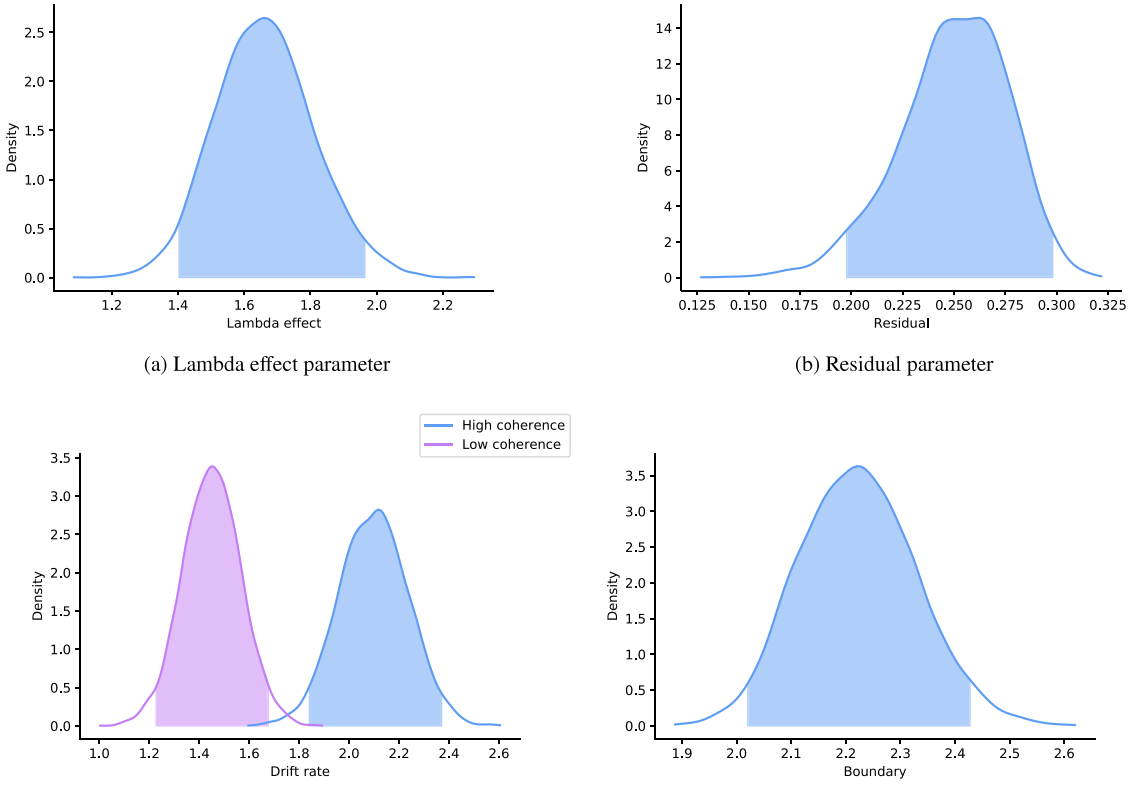


(h) Hierarchical residual distribution of trial-to-trial variability across spatial attention

Fig. 11. Posterior distributions of Hierarchical neuro-cognitive Bayesian *Model 4A Full*. The two figures in the first row correspond to the four conditions: prioritized and high coherence (PHC), not-prioritized and high coherence (NPHC), prioritized and low coherence (PLC), and not-prioritized and low coherence (NPLC). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of possible hypotheses for neuro-cognitive models. At the level of behavioral data, the result of the model comparison shows that the boundary and drift rate parameters are not mainly manipulated by valid top-down cues, see the section "Testing the hypotheses of the behavioral data" and Table 5 in Appendix.

The results of model comparison among the hierarchical neuro-cognitive models show that *Models 1B*, *2B*, *3B*, and *4B*, which assume that spatial cueing can shift boundary parameters, were not selected as the best models, see Table 1. The table also shows that *Models 1C*, *2C*, *3C*, and *4C*, which assume that spatial cueing can manipulate drift rate, were not selected as the best models. Although WAIC for *Model 4C* is similar to *Model 4A*, -ELPD of *Model 4C* is larger. In line with the parsimony principle, we



(a) Lambda effect parameter

(b) Residual parameter

(c) evidence accumulation rate across two high and low coherence

(d) Boundary parameter

Fig. 12. . Parameter posteriors of *Model 1A* and simulated participant data 1. The shaded area represents the 95% Bayesian credible interval (BCI) of posteriors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chose a simple model over the complex model, so *Model 4A* is the best model.

When comparing the best model, *Model 4A*, and *Model 4A Full*, we can conclude that parameter of boundary and drift rate was not mainly shifted by valid cues. By comparing posterior distributions of parameters across two models, we can see whether we would draw the same conclusions To do so, compare Figs. 6(b) and 6(c) to Figs. 11(e) and 11(f). The posteriors of the residual (r) and the coefficient parameter (λ) in *Model 4A* are similar in *Model 4A Full* that lets the boundary and drift rate change. Therefore we would have drawn the same conclusions even if this model best described the data.

4.3. Individual differences in the findings of this study

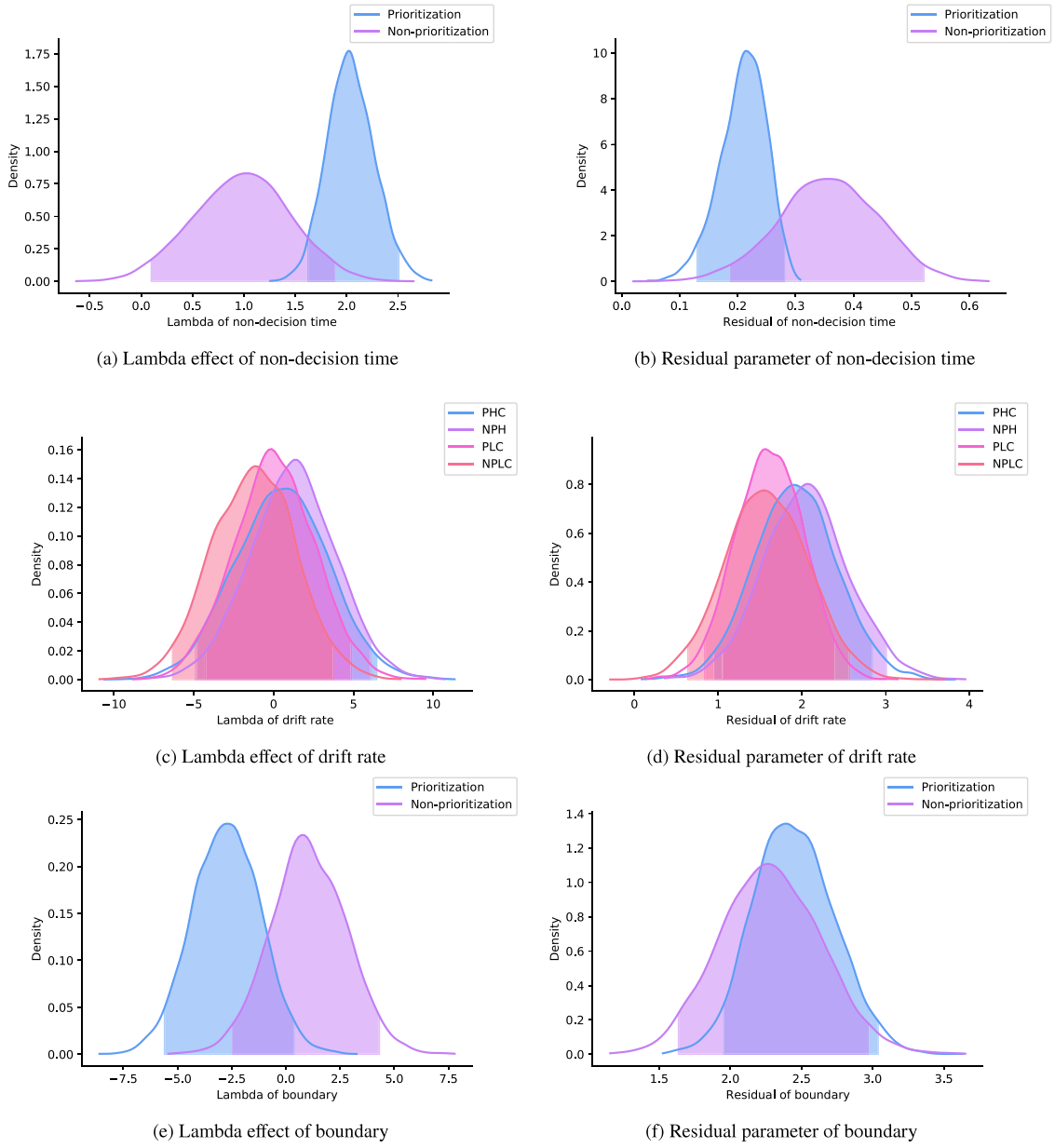
While over all participants it was true that model results suggest both VET and other non-decision time components are affected by spatial cueing, the exact results varied by participants. We suspect that these could be due to actual strategic or cognitive differences across individuals. Or the results could be due to different levels of noise and contaminants in the EEG N200 measures and behavioral data.

Specifically, we found that most participants' data were best described by *Model 3A* while other participants' data were best described by *Model 4A* (see Table 4). This implies that for most participants, only the relationship of N200 to VET is affected by spatial prioritization. However, when evaluating a particular participant's data with less noisy and seemingly more precise single-trial N200 estimates, it was found that *Model 4A* best describes the data. Further study is necessary to determine if there are true cognitive differences in spatial attention between individuals or if these results were driven by EEG artifact differences between the two conditions in some participants.

4.4. Limitations of this study and future work

We decided to focus on occipito–parietal electrodes based on our literature review and previous research. We used the first or second component of SVD to extract single-trial latencies which are assumed than they can track single-trial visual encoding. The SVD can provide a weighted map for all electrodes and waveforms at the same time such that the weighted map should be positive and the waveform should have a negative peak around occipito–parietal sensors. However, the use of advanced signal processing techniques for extracting single-trial measures of the EEG activities could be developed to improve inference with less trial noise or artifact. Furthermore, neuro-cognitive models that better account for sources of artifact in single trials could be fit to EEG and behavioral data jointly.

Weindel et al. have proposed a novel approach based on the onsets of electromyographical (EMG) activity to separate response time into a pre-motor (PMT) and a motor time (MT) in perceptual decision tasks (Weindel et al., 2021; Weindel, Gajdos, Burle, & Alario, 2022). Furthermore, researchers have shown that such data can be described well by Dual-Threshold Diffusion Models (DTDMs) (Servant, Logan, Gajdos, & Evans, 2021). In the future researchers should seek to combine both EEG and EMG activity within neuro-cognitive models to better decompose non-decision components from decision components in perceptual decision making tasks. These models will provide even more inference into the role of spatial attention on non-decision time components. Specifically, these models can then whether evidence accumulation models that include non-decision time stages between evidence accumulation stages better explain spatial prioritization effects on non-decision times. This is because VET, evidence accumulation, and MT could be separately identified within models. These models can also directly test whether spatial prioritization affects motor time (MT) with appropriate experimental



(a) Lambda effect of non-decision time

(b) Residual parameter of non-decision time

(c) Lambda effect of drift rate

(d) Residual parameter of drift rate

(e) Lambda effect of boundary

(f) Residual parameter of boundary

Fig. 13. . Parameter posteriors of *Model 4A Full* and simulated participant data 1. The shaded area represents the 95% Bayesian credible interval (BCI) of posteriors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

paradigms. This was not a question that we could easily answer with existing datasets.

5. Conclusion

Top-down spatial prioritization amplifies individuals' performance to make quicker decisions since they do not require to take some time to search the visual field to find the location of the stimulus appearance. Neuro-cognitive modeling incorporated human behavior and electrophysiological activity at the same time to constrain each in order to decompose the non-decision time and understand differential effects of spatial attention on visual encoding time (VET) and other non-decision time processes.

We revealed that both visual stimulus encoding and other non-VET non-decision processes are manipulated by spatial cueing. The group-level effect of single-trial N200 latencies on non-decision times was positive and group-level residual time for prioritization was lower than non-prioritization. Moreover, we show that occipital–parietal deflections extracted by SVD play a key role in the study of spatial attention, derived directly from N200 latencies thought to track visual encoding time and the onset of evidence accumulation (Loughnane et al., 2016; Nunez et al., 2019).

Funding

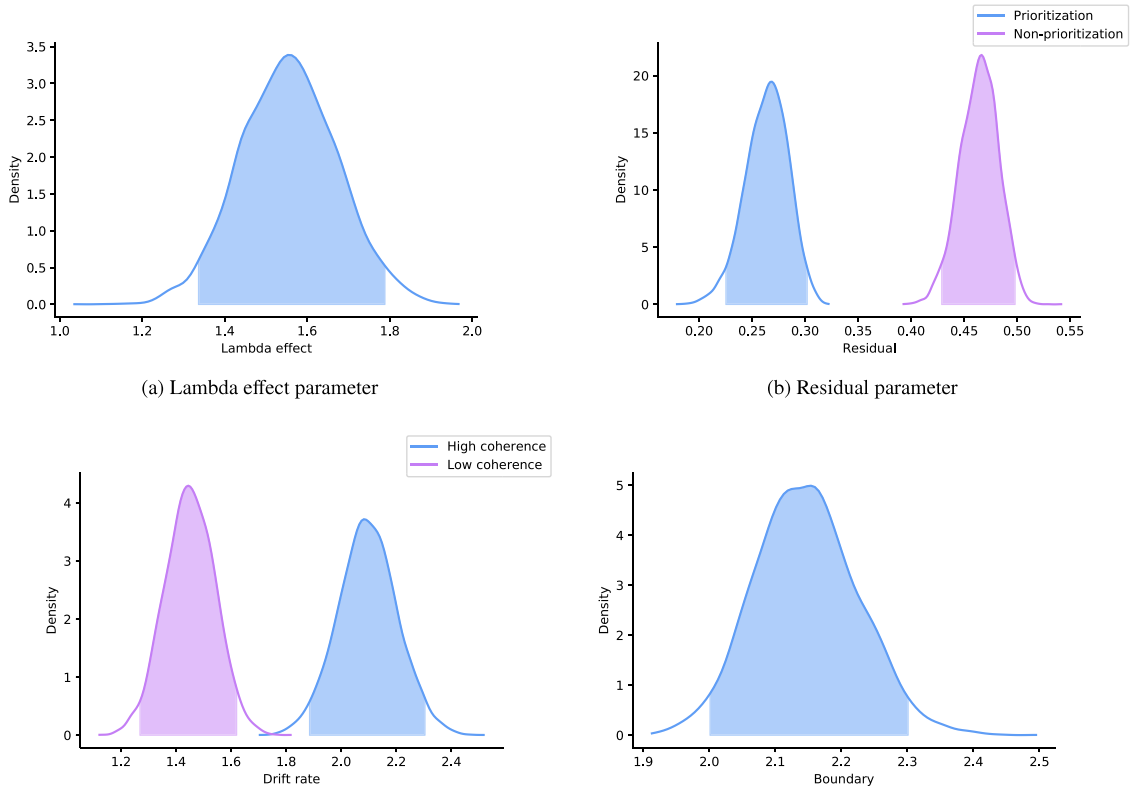
JAR was supported in part by IRAN Cognitive Sciences & Technologies Council (Grant No. 9311), and Iran National Science Foundation, INSF (Grant No. 99010447 and 4003215).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.



(a) Lambda effect parameter

(b) Residual parameter

(c) evidence accumulation rate across two high and low coherence

(d) Boundary parameter

Fig. 14. . Parameter posteriors of *Model 2A* and simulated participant data 2. The shaded area represents the 95% Bayesian credible interval (BCI) of posteriors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix A. Additional experimental details

To produce face and car images with two different phase coherences, the phase spectrum of the original images was combined with a phase spectrum of an image of uniformly distributed random noise using the weighted mean phase (WMP) method as follows (Dakin, Hess, Ledgeway, & Achtman, 2002; Ostwald et al., 2012; Georgie et al., 2018):

$$\phi_s = \begin{cases} \arctan(S_\phi/C_\phi) & C_\phi > 0 \\ \arctan(S_\phi/C_\phi) + \pi & C_\phi < 0, S_\phi > 0 \\ \arctan(S_\phi/C_\phi) - \pi & C_\phi < 0, S_\phi < 0 \end{cases} \tag{11a}$$

$$S_\phi = w\sin(\phi_o) + (1-w)\sin(\phi_n) \tag{11b}$$

$$C_\phi = w\cos(\phi_o) + (1-w)\cos(\phi_n) \tag{11c}$$

where ϕ_s is the phase spectrum of new images, ϕ_o is the phase spectrum of the original images, ϕ_n is the phase spectrum of a noise image, w is in the interval $[0, 1]$ which represents the signal-to-noise ratio of images. The new scrambled stimuli were constructed by two different high ($w = 0.9$) and low ($w = 0.5$) phase coherence.

Appendix B. Testing hypotheses on behavioral data

We re-analyzed this dataset with bootstrapping to enlarge the dataset to further verify our results in Ghaderi-Kangavari et al. (2021), these results are reported here. We proposed four hierarchical models containing, $model_t$ assuming spatial attention may shift non-decision time, $model_v$ assuming spatial attention may shift drift rate parameter, $model_a$ assuming spatial attention may manipulate boundary, and $model_p$ assuming spatial attention manipulates no parameters significantly. Based on model selection criteria for Hierarchical Bayesian model, $model_t$ is better model for all bootstraps (see Figs. 9 and 8)

Table 5
This is the result of bootstrapping over only behavioral data. We used 30 times bootstraps and for each bootstrap, we chose 24 participants from the original participants by random and replacement. We tested four models in which of main DDM's parameters are manipulated by a top-down cue. In all modes, we fixed the effect of the coherence independent variable (coher) to the drift rate parameter (δ). For the random variable of spatial cueing, we can test which one of the drift rate, non-decision time parameter (τ), and boundary separation (α) is changed. Based on four criteria model fitting, $Model_t$ is a better model to describe the manipulation of spatial prioritization. Lower WAICs and -ELPD indicate better fits to data after accounting for model complexity. PWAIC shows the effective number of parameters and the complexity of models.

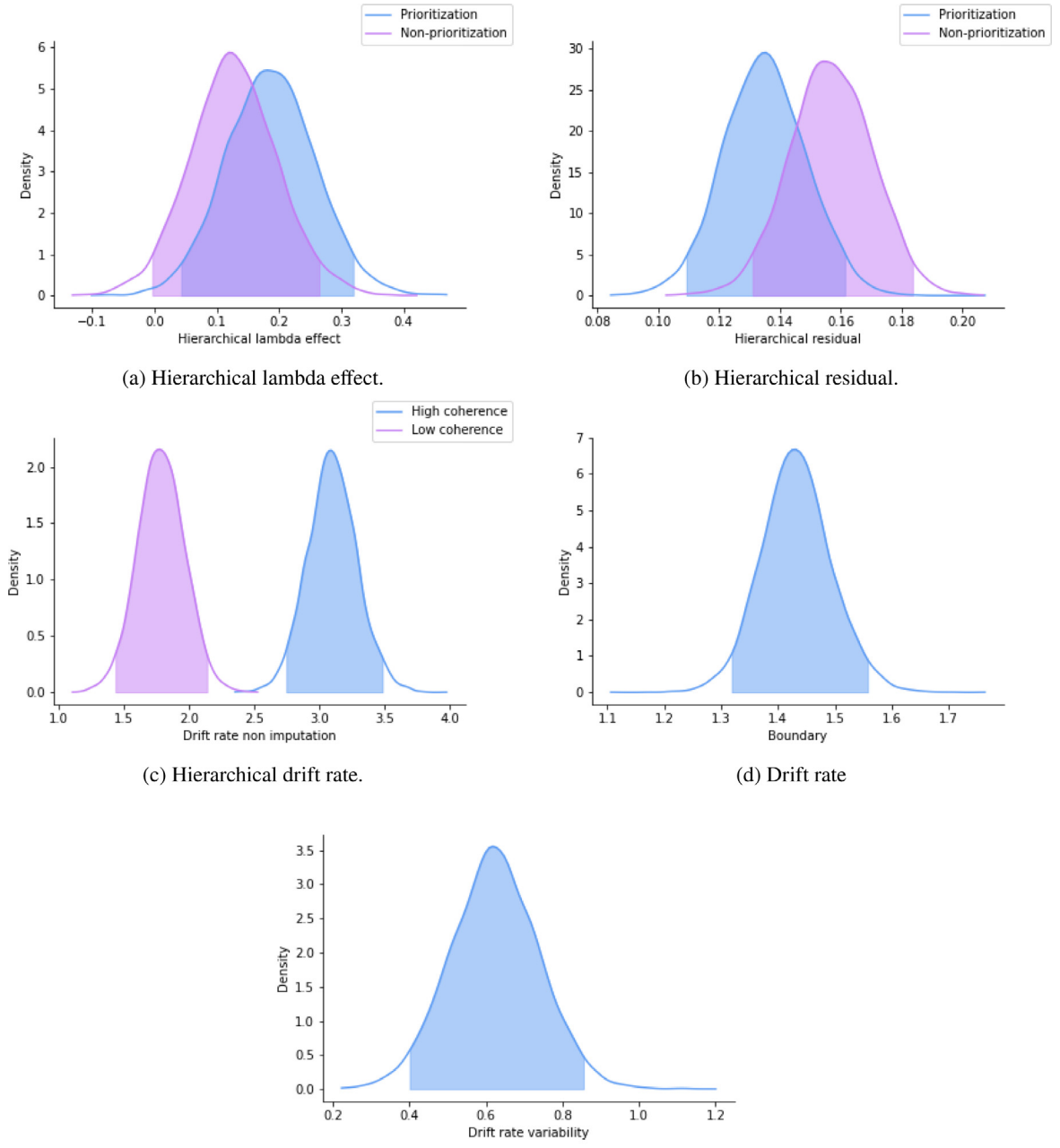
	$\delta \sim coher$	$\delta \sim spat$	$\tau \sim spat$	$\alpha \sim spat$	WAIC	PWAIC	-ELPD
$Model_p$	√				−3983	134	−1983
$Model_v$	√	√			−4039	170	−2011
$Model_t$	√		√		**−4125**	163	**−2050**
$Model_a$	√			√	−4045	154	−2017

For $Model_t$ (Fig. 9(a)): indices of i refer to trials that do not vary with random variables. Also, indices of j participants, k conditions, k_1 coherence and k_2 spatial prioritization were set. Distributions of parameters for each participant i and condition k are as follows:

$$\left(\delta_{jk_1}|\mu_{(\delta)k_1}, \sigma_{(\delta)}\right) \sim \mathcal{N}(\mu_{(\delta)k_1}, \sigma^2_{(\delta)}) \in (0, 6), \mu_{(\delta)k_1} \sim \mathcal{N}(2, 3^2), \sigma_{(\delta)} \sim \Gamma(1, 1). \tag{12a}$$

$$\left(\tau_{jk_2}|\mu_{(\tau)k_2}, \sigma_{(\tau)}\right) \sim \mathcal{N}(\mu_{(\tau)k_2}, \sigma^2_{(\tau)}) \in (0, 1), \mu_{(\tau)k_2} \sim \mathcal{N}(.3, .2^2), \sigma_{(\tau)} \sim \Gamma(.3, 1). \tag{12b}$$

$$\left(\alpha_j|\mu_{(\alpha)}, \sigma_{(\alpha)}\right) \sim \mathcal{N}(\mu_{(\alpha)}, \sigma^2_{(\alpha)}) \in (0, 3), \mu_{(\alpha)} \sim \mathcal{N}(1, 2^2), \sigma_{(\alpha)} \sim \Gamma(1, 1). \tag{12c}$$



(a) Hierarchical lambda effect.

(b) Hierarchical residual.

(c) Hierarchical drift rate.

(d) Drift rate

(e) Hierarchical trial-to-trial drift rate variability

Fig. 15. Posterior distributions of Hierarchical Bayesian *Model 4A* for non-imputed N200 latencies.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\left(\eta_j|\mu_{(\eta)}, \sigma_{(\eta)}\right) \sim \mathcal{N}(\mu_{(\eta)}, \sigma^2_{(\eta)}) \in (0, 3), \mu_{(\eta)} \sim \mathcal{N}(1, .5^2),$$
$$\sigma_{(\eta)} \sim \Gamma(1, 1). \tag{12d}$$
$$y_{ijk} \sim \mathit{Wiener}\left(\alpha_j, \tau_{jk_1}, \beta, \delta_{jk_2}, \eta_j\right). \tag{12e}$$

For Model_v (Fig. 9(b)): prior distributions of drift-rate and non-decision time parameters were changed as follows:

$$\left(\delta_{jk}|\mu_{(\delta)k}, \sigma_{(\delta)}\right) \sim \mathcal{N}(\mu_{(\delta)k}, \sigma^2_{(\delta)}) \in (0, 6), \mu_{(\delta)k} \sim \mathcal{N}(2, 3^2),$$
$$\sigma_{(\delta)} \sim \Gamma(1, 1). \tag{13a}$$
$$\left(\tau_j|\mu_{(\tau)}, \sigma_{(\tau)}\right) \sim \mathcal{N}(\mu_{(\tau)}, \sigma^2_{(\tau)}) \in (0, 1).\mu_{(\tau)} \sim \mathcal{N}(.3, .2^2),$$
$$\sigma_{(\tau)} \sim \Gamma(.3, 1). \tag{13b}$$

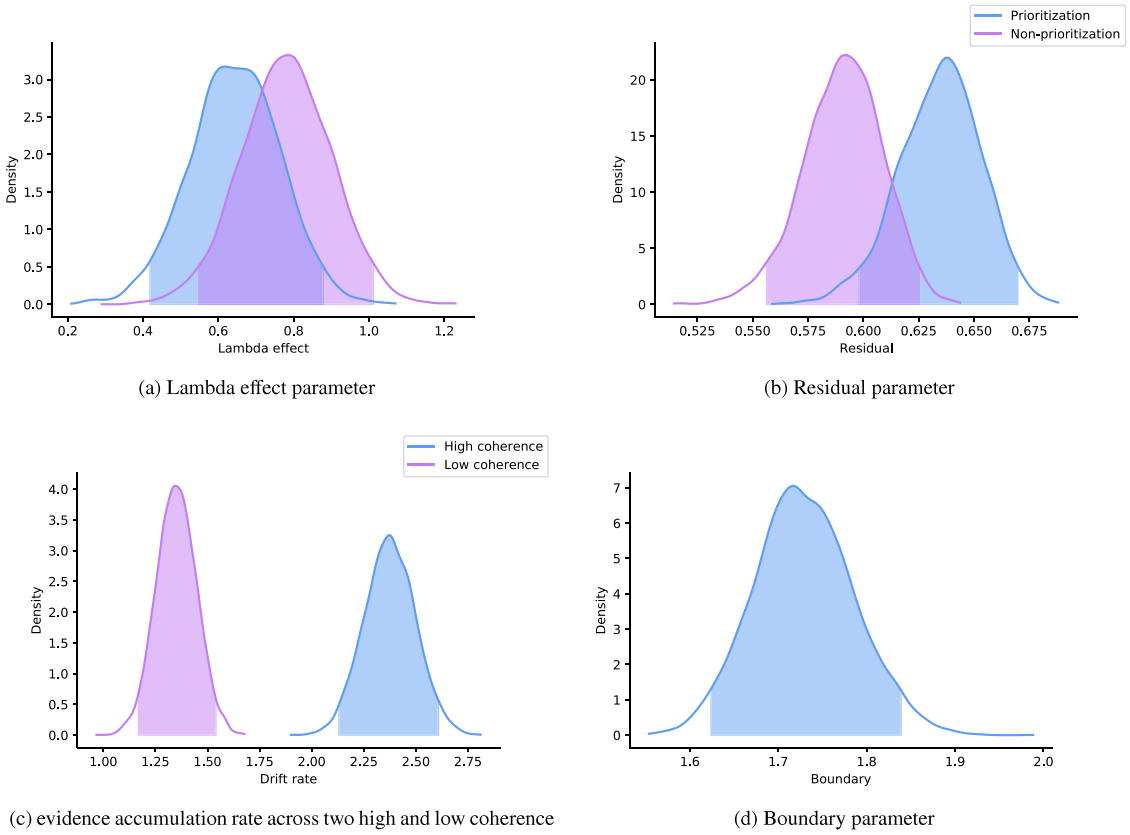
For Model_a (Fig. 9(c)): prior distributions of drift-rate free from coherence and non-decision time parameters free from spatial prioritization as follows:

$$\left(\delta_{jk_1}|\mu_{(\delta)k_1}, \sigma_{(\delta)}\right) \sim \mathcal{N}(\mu_{(\delta)k_1}, \sigma^2_{(\delta)}) \in (0, 6), \mu_{(\delta)k_1}$$
$$\sim \mathcal{N}(2, 3^2), \sigma_{(\delta)} \sim \Gamma(1, 1). \tag{14a}$$
$$\left(\alpha_{jk_2}|\mu_{(\tau)k_2}, \sigma_{(\alpha)}\right) \sim \mathcal{N}(\mu_{(\alpha)k_2}, \sigma^2_{(\alpha)}) \in (0, 1), \mu_{(\alpha)k_2}$$
$$\sim \mathcal{N}(.3, .2^2), \sigma_{(\alpha)} \sim \Gamma(.3, 1). \tag{14b}$$

For Model_p (Fig. 9(d)): only prior distributions of drift-rate free from coherence as follows:

$$\left(\delta_{jk_1}|\mu_{(\delta)k_1}, \sigma_{(\delta)}\right) \sim \mathcal{N}(\mu_{(\delta)k_1}, \sigma^2_{(\delta)}) \in (0, 6), \mu_{(\delta)k_1}$$
$$\sim \mathcal{N}(2, 3^2), \sigma_{(\delta)} \sim \Gamma(1, 1). \tag{15a}$$

Finally, R-hat was less than 1.1 for all parameters in each model and bootstrap. Also, The mean effective sample size for all parameters across bootstraps and models was 8200.



(a) Lambda effect parameter

(b) Residual parameter

(c) evidence accumulation rate across two high and low coherence

(d) Boundary parameter

Fig. 16. Parameter posteriors of *Model 4A* fit Simulation 4. The shaded area represents the 95% Bayesian credible interval (BCI) of posteriors.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Pearson correlation coefficients and p-values with associated of single-trial N200 latencies and response times for each participant. Note that outlier single-trial N200 latencies on the window boundaries were removed before calculating the correlation coefficients.

Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r	0.006	−0.019	0.088	0.078	0.024	−0.07	0.086	0.226	−0.082	0.214	0.051	0.093	0.033	0.073	0.095
p-value	0.931	0.759	0.163	0.218	0.683	0.291	0.187	0.0003	0.200	0.002	0.452	0.156	0.617	0.287	0.136

Table 7
The results of using Bayesian and frequentist inference to understand the differences of the mean N200 latencies between the non-prioritized and prioritized conditions, $z_{(x)np} - z_{(x)p}$. The first row contains the participant number. The second row contains the Bayesian probability (e.g. evidence in the Bayesian framework) that the difference is larger than 0. Specifically, *Pr* is the integral of the posterior distribution of $z_{(x)np} - z_{(x)p}$ from 0 to ∞ for each participant. Third row shows the *p*-value of frequentist t-test between two conditions for each participant. Only one participant (participant 5) had a significant difference in mean N200 latencies between the two conditions with a Type 1 error rate of 0.01 (indicated by the row in bold). Bayesian inference indicates that there is a 99% probability of the difference being greater than 0 for this participant.

Participants	1	2	3	4	**5**	6	7	8	9	10	11	12	13	14	15
Pr	0.70	0.78	0.49	0.86	**0.99**	0.64	0.44	0.75	0.23	0.93	0.18	0.34	0.94	0.67	0.32
p-value	0.55	0.38	0.95	0.30	**0.008**	0.83	0.82	0.57	0.36	0.16	0.30	0.63	0.14	0.65	0.53

Table 8
The results of model comparison for hierarchical neuro-cognitive models of behavior and N200 latencies in which the outlier values were thrown-out instead of imputed. Lower WAIC, -LPPD, and -ELPD indicate better fits to data.

	$\delta \sim coher$	$r \sim spat$	$\lambda \sim spat$	$all \sim spat$	WAIC	-LPPD	PWAIC	-ELPD
Model 1A	√				−2095	−1139	92	−1041
Model 2A	√	√			−2126	−1168	105	−1055
Model 3A	√		√		−2134	−1167	100	−1059
Model 4A	√	√	√		**−2135**	−1176	108	**−1060**
Model 4A Full	√			√	−2125	**−1212**	150	−1053

Appendix C. The full neuro-cognitive model, *Model 4A Full*

Model 4A Full with all variables free to vary by trial-to-trial N200 latencies:

$$\left(r_{(\delta)jk}|\mu_{(r,\delta)k},\sigma_{(r,\delta)}\right)\sim\mathcal{N}\left(\mu_{(r,\delta)k},\sigma^2_{(r,\delta)}\right),\ \mu_{(r,\delta)k}\sim\mathcal{N}(2,4^2),$$
$$\sigma_{(r,\delta)}\sim\Gamma(1,1). \tag{16a}$$
$$\left(\lambda_{(\delta)jk}|\mu_{(\lambda,\delta)k},\sigma_{(\lambda,\delta)}\right)\sim\mathcal{N}\left(\mu_{(\lambda,\delta)k},\sigma^2_{(\lambda,\delta)}\right),\ \mu_{(\lambda,\delta)k}\sim\mathcal{N}(3,5^2),$$
$$\sigma_{(\lambda,\delta)}\sim\Gamma(1,1).$$
$$\left(r_{(\alpha)jk_2}|\mu_{(r,\alpha)k_2},\sigma_{(r,\alpha)}\right)\sim\mathcal{N}\left(\mu_{(r,\alpha)k_2},\sigma^2_{(r,\alpha)}\right),\ \mu_{(r,\alpha)k_2}\sim\mathcal{N}(1,3^2),$$
$$\sigma_{(r,\alpha)}\sim\Gamma(1,1) \tag{16b}$$
$$\left(\lambda_{(\alpha)jk_2}|\mu_{(\lambda,\alpha)k_2},\sigma_{(\lambda,\alpha)}\right)\sim\mathcal{N}\left(\mu_{(\lambda,\alpha)k_2},\sigma^2_{(\lambda,\alpha)}\right),\ \mu_{(\lambda,\alpha)k_2}\sim\mathcal{N}(3,5^2),$$
$$\sigma_{(\lambda,\alpha)}\sim\Gamma(1,1). \tag{16c}$$
$$\left(r_{(\tau)jk_2}|\mu_{(r,\tau)k_2},\sigma_{(r,\tau)}\right)\sim\mathcal{N}\left(\mu_{(r,\tau)k_2},\sigma^2_{(r,\tau)}\right),$$
$$\mu_{(r,\tau)k_2}\sim\mathcal{N}(.2,.4^2),\sigma_{(r,\tau)}\sim\Gamma(.1,1). \tag{16d}$$
$$\left(\lambda_{(\tau)jk_2}|\mu_{(\lambda,\tau)k_2},\sigma_{(\lambda,\tau)}\right)\sim\mathcal{N}\left(\mu_{(\lambda,\tau)k_2},\sigma^2_{(\lambda,\tau)}\right),$$
$$\mu_{(\lambda,\tau)k_2}\sim\mathcal{N}(.5,2^2),\sigma_{(\lambda,\tau)}\sim\Gamma(.1,1). \tag{16e}$$
$$\left(r_{(\eta)jk_2}|\mu_{(r,\eta)k_2},\sigma_{(r,\eta)}\right)\sim\mathcal{N}\left(\mu_{(r,\eta)k_2},\sigma^2_{(r,\eta)}\right)\in(0,),$$
$$\mu_{(r,\eta)k_2}\sim\mathcal{N}(1,.5^2),\sigma_{(r,\eta)}\sim\Gamma(1,1). \tag{16f}$$
$$\left(\lambda_{(\eta)jk_2}|\mu_{(\lambda,\eta)k_2},\sigma_{(\lambda,\eta)}\right)\sim\mathcal{N}\left(\mu_{(\lambda,\eta)k_2},\sigma^2_{(\lambda,\eta)}\right)\in(0,),$$
$$\mu_{(\lambda,\eta)k_2}\sim\mathcal{N}(2,4^2),\sigma_{(\lambda,\eta)}\sim\Gamma(1,1). \tag{16g}$$
$$\left(\mu_{(z)jk}|\mu_{(z)k},\sigma_{(z)}\right)\sim\mathcal{N}\left(\mu_{(z)k},\sigma^2_{(z)}\right)\in(0,.4),\ \mu_{(z)k}$$
$$\sim\mathcal{N}(.15,.1^2),\sigma_{(z)},\sim\Gamma(.1,1). \tag{16h}$$
$$y_{ijk}\sim Wiener\Big(r_{(\alpha)jk_2}+\lambda_{(\alpha)jk_2}x_{ijk},r_{(\tau)jk_2}+\lambda_{(\tau)jk_2}x_{ijk},\beta,$$
$$r_{(\delta)jk}+\lambda_{(\delta)jk}x_{ijk},r_{(\eta)jk_2}+\lambda_{(\eta)jk_2}x_{ijk}\Big). \tag{16i}$$
$$x_{ijk}\sim\mathcal{N}\left(z_{(x)jk},\sigma^2_{(x)}\right)\in(.101,.248),\sigma_{(x)}\sim\Gamma(.1,1). \tag{16j}$$

C.1. Hierarchical neuro-cognitive models without imputed N200 latencies

See Table 8.

References

Barnard, J., & Rubin, D. B. (1999). Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955.

Bolam, J., Boyle, S. C., Ince, R. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. *NeuroImage*, *247*, Article 118841.

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), Article e13049.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nature Neuroscience*, *2*(4), 370–374.

Bridwell, D. A., Cavanagh, J. F., Collins, A. G., Nunez, M. D., Srinivasan, R., Stober, S., et al. (2018). Moving beyond ERP components: a selective review of approaches to integrate EEG and behavior. *Frontiers in Human Neuroscience*, *12*, 106.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Chica, A. B., Bartolomeo, P., & Lupiáñez, J. (2013). Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural Brain Research*, *237*, 107–123.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571.

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, *56*(11), Article e13437.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.

Cohen, M. X. (2021). *Linear algebra: theory, intuition, code*. [Sl]: Sincxpress Bv.

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

Dakin, S., Hess, R., Ledgeway, T., & Achtman, R. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, *12*(14), R476–R477.

Dmochowski, J. P., & Norcia, A. M. (2015). Cortical components of reaction-time during perceptual decisions in humans. *PLoS One*, *10*(11), Article e0143339.

Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E.-J., & Brown, S. D. (2017). The computations that support simple decision-making: A comparison between the diffusion and urgency-gating models. *Scientific Reports*, *7*(1), 1–13.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–170.

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review*, *26*(4), 1099–1121.

Frank, M. J. (2015). Linking across levels of computation in model-based cognitive neuroscience. In *An introduction to model-based cognitive neuroscience* (pp. 159–177). Springer.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain monte carlo: stochastic simulation for bayesian inference*. CRC Press.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *96*(6), 3314–3319.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd). Boca Raton, FL: Taylor & Francis Group, LLC.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Georgie, Y. K., Porcaro, C., Mayhew, S. D., Bagshaw, A. P., & Ostwald, D. (2018). A perceptual decision making EEG/fMRI data set. Article 253047, BioRxiv.

Ghaderi-Kangavari, A., Rad, J. A., & Nunez, M. D. (2022). A general integrative neurocognitive modeling framework to jointly describe EEG and decision-making on single trials.

Ghaderi-Kangavari, A., Rad, J. A., Parand, K., Ebrahimpour, R., & Nunez, M. D. (2021). How spatial attention affects the decision process: looking through the lens of Bayesian hierarchical diffusion model & EEG analysis. BioRxiv.

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

Gomez, P., Ratcliff, R., & Childers, R. (2015). Pointing, looking at, and pressing keys: A diffusion model account of response modality. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1515.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.

Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306.

Hall, J. E., & Hall, M. E. (2020). *Guyton and hall textbook of medical physiology e-book*. Elsevier Health Sciences.

Hunt, A. R., Reuther, J., Hilchey, M. D., & Klein, R. M. (2019). The relationship between spatial attention and eye movements. *Processes of Visuospatial Attention and Working Memory*, *41*, 255–278.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634.

Imani, E., Harati, A., Pourreza, H., & Goudarzi, M. M. (2021). Brain-behavior relationships in the perceptual decision-making process through cognitive processing stages. *Neuropsychologia*, *155*, Article 107821.

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, *225*, Article 117465.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764.

Klein, R. M. (1994). Perceptual-motor expectancies interact with covert visual orienting under conditions of endogenous but not exogenous control. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie ExpÉRimentale*, *48*(2), 167.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*, 193.
Lamme, V. A., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*(7), 1044–1053.
Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. In *Special Issue on Hierarchical Bayesian Models*: *Journal of Mathematical Psychology*, [ISSN: 0022-2496] In *Special Issue on Hierarchical Bayesian Models*: *55*(1).1–7,
Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.
Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., & O'Connell, R. G. (2016). Target selection signals influence perceptual decisions by modulating the onset and rate of evidence accumulation. *Current Biology*, *26*(4), 496–502.
Luck, S. J. (2005). An introduction to event related potentials and their neural origins. *An introduction to the event related potential technique, vol. 11*. The MIT Press.
Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT Press.
Macaluso, E., & Driver, J. (2001). Spatial attention and crossmodal interactions between vision and touch. *Neuropsychologia*, *39*(12), 1304–1316.
Molenberghs, P., Mesulam, M. M., Peeters, R., & Vandenberghe, R. R. (2007). Remapping attentional priorities: differential contribution of superior parietal lobule and intraparietal sulcus. *Cerebral Cortex*, *17*(11), 2703–2712.
Moore, D., McCabe, G., Duckworth, W., & Sclove, S. (2003). Bootstrap methods and permutation tests. In *The practice of business statistics: using data for decisions*. New York: WH Freeman.
Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.
Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage*, *197*, 93–108.
Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology*, *76*, 117–130.
Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2022). A tutorial on fitting joint models of M/EEG and behavior to understand cognition. PsyArXiv.
Ostwald, D., Porcaro, C., Mayhew, S. D., & Bagshaw, A. P. (2012). EEG-fMRI based information theoretic characterization of the human perceptual decision system. *PLoS One*, *7*(4), Article e33896.
Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., & Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal of Mathematical Psychology*, *84*, 20–48.
Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1.
Palmeri, T. J., Love, B. C., & Turner, B. M. (2017). Model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 59–64.
Philiastides, M., Diaz, J., & Gherman, S. (2017). Spatiotemporal characteristics and modulators of perceptual decision-making in the human brain. In *Decision Neuroscience* (pp. 137–147). Elsevier.
Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience*, *26*(35), 8965–8975.
Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, *69*(10), 1864–1875.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, [ISSN: 1364-6613] *20*(4), 260–281.
Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*(21), 9475–9489.
Sagar, V., Sengupta, R., & Sridharan, D. (2019). Dissociable sensitivity and bias mechanisms mediate behavioral effects of exogenous attention. *Scientific Reports*, *9*(1), 1–13.
Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*(11), 2917–2926.
Schubert, A.-L., Nunez, M. D., Hagemann, D., & Vandekerckhove, J. (2019). Individual differences in cortical processing speed predict cognitive abilities: A model-based cognitive neuroscience account. *Computational Brain & Behavior*, *2*(2), 64–84.
Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology: General*, *150*(12), 2435–2454.
Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking theoretical decision-making mechanisms in the simon task with electrophysiological data: A model-based neuroscience study in humans. *Journal of Cognitive Neuroscience*, *28*(10), 1501–1521.
Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791–806.
Sheliga, B., Riggio, L., & Rizzolatti, G. (1995). Spatial attention and eye movements. *Experimental Brain Research*, *105*(2), 261–275.
Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.
Simon, S. R., Meunier, M., Piettre, L., Berardi, A. M., Segebarth, C. M., & Boussaoud, D. (2002). Spatial attention and memory versus motor preparation: premotor cortex involvement as revealed by fMRI. *Journal of Neurophysiology*, *88*(4), 2047–2057.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *64*(4), 583–639.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.
Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, *26*(37), 9494–9502.
Tillman, G., Van Zandt, T., & Logan, G. D. (2020). Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making. *Psychonomic Bulletin & Review*, *27*(5), 911–936.
Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79.
Turner, B. M., Forstmann, B. U., Steyvers, M., et al. (2019). *Joint models of neural and behavioral data*. Springer.
Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, *72*, 193–206.
Turner, B. M., Van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: the neural drift diffusion model.. *Psychological Review*, *122*(2), 312.
Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). 14 Model comparison and the principle of parsimony. In *The oxford handbook of computational and* (p. 300).
VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.
Verdonck, S., Loossens, T., & Philiastides, M. G. (2021). The leaky integrating threshold and its impact on evidence accumulation models of choice response time (RT). *Psychological Review*, *128*(2), 203.
Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: a tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, *46*(1), 15–28.
Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*(1), 28–50.
Weindel, G., Anders, R., Alario, F., Burle, B., et al. (2021). Assessing model-based inferences in decision making with single-trial response time decomposition.. *Journal of Experimental Psychology: General*, *150*(8), 1528.
Weindel, G., Gajdos, T., Burle, B., & Alario, F.-X. (2022). The decisive role of non-decision time for interpreting decision making models. PsyArXiv.
Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 14.
Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. *Journal of Neuroscience*, *20*(6), RC63.
Wright, D. B., London, K., & Field, A. P. (2011). Using bootstrap estimation and the plug-in principle for clinical psychology data. *Journal of Experimental Psychopathology*, *2*(2), 252–270.
Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306.