

Quantifying and Characterizing Dynamic Mechanisms of Cognitive Control

Emily Ruth Weichart
Mayfield Heights, Ohio

M.A., The Ohio State University, 2017
B.S., The Ohio State University, 2013

A Dissertation presented to the Graduate Faculty of the University of Virginia
in Candidacy for the Degree of Doctor of Philosophy

Department of Psychology

University of Virginia
August, 2020

Per B. Sederberg, Ph.D.
Chad S. Dodson, Ph.D.
T. Jason Druzgal, M.D., Ph.D.
Nicole M. Long, Ph.D.
C. Daniel Meliza, Ph.D.

Abstract

The human cognitive system is remarkably flexible and can rapidly adjust to shifting task demands and perceptual inputs. This flexibility is achieved through cognitive control, which comprises the goal-directed modulation of attention and behavior. Cognitive control is often measured by proxy, via speed and accuracy on behavioral tasks that are thought to access the brain networks of interest. However, there is much that remains unknown about how task-related neural subprocesses—visual perception, attention allocation, cognitive control, decision making—work together to produce behavior. My dissertation work uses computational modeling approaches to decompose behavior into its constituent mechanisms, with the overarching goals of understanding behaviors from the vantage point of neural mechanisms and quantifying those mechanisms in a meaningful way. I begin by introducing the flanker task as a viable means of tracking cognitive control on immediate timescales via a deep brain stimulation clinical case study. Next, I present competing theories for how and when cognitive control engages on a within-trial basis, and the mathematical details of generative models that instantiate each theory in a testable way. Using the best-fitting model across three experiments, I further investigate individual differences in the allocation of spatial attention. Finally, I package my findings into a tool for measuring and comparing mechanisms associated with cognitive control within- and between subjects, and demonstrate Bayesian-inspired methods for identifying meaningful cognitive changes through time.

To LEGO Ron, for all the laughs during the humid years

Acknowledgements

I first would like to thank Per Sederberg for ten years of guidance, support, and boundless enthusiasm. It is slightly mortifying to think about how much Per has personally taken the time to teach me through the years, from the basics of coding and analyses in my undergrad days to how to review the author proofs of my first publication. I am unbelievably lucky to have wandered into the lab of someone who is both a powerhouse of a researcher and a genuinely caring human being, and I couldn't have asked for a more wonderful mentor. Thank you for your steadfast faith in me, despite my constant insistence that everything I do is a nightmare.

I next want to thank Brandon Turner, a cherished friend and collaborator, for lending his invaluable insights and expertise to my projects through the years and for being a calming and pleasant presence in my academic life. You've been an inspiration, both for your impossible work ethic and your uncommonly kind heart.

Thank you to my dearest friend and mysterious wind, Zachary Niese, for somehow being the cool logic that I so often needed while also being the dazzling absurdity that I sometimes needed more. Speaking fluent nonsense with you has been the joy of my life.

Finally, I would like to thank my mom, Virginia Weichart, who worked so hard to give me every opportunity to learn, ask questions, and follow my interests. It is because of her early efforts and continued encouragement that graduate school was ever a possibility for me. I hope I can one day be even half as wise as she was.

Table of Contents

| | |
|--|------------|
| ABSTRACT | I |
| DEDICATION | II |
| ACKNOWLEDGEMENTS | III |
| TABLE OF CONTENTS | IV |
| LIST OF TABLES | IX |
| LIST OF FIGURES | X |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1 MOTIVATION..... | 1 |
| 1.2 ORGANIZATION | 3 |
| CHAPTER 2: COGNITIVE TASK PERFORMANCE DURING TITRATION PREDICTS DEEP BRAIN STIMULATION TREATMENT EFFICACY | 7 |
| 2.1 INTRODUCTION..... | 7 |
| 2.2 MATERIALS AND METHODS | 10 |
| 2.2.1 <i>Participants</i> | 10 |
| 2.2.2 <i>Surgery</i> | 11 |
| 2.2.3 <i>Titration</i> | 12 |

| | | |
|--|---|-----------|
| 2.2.4 | <i>Long-Term Monitoring</i> | 13 |
| 2.2.5 | <i>EEG Recording</i> | 14 |
| 2.3 | RESULTS | 16 |
| 2.3.1 | <i>Optimal Stimulation Settings for Weight Loss</i> | 16 |
| 2.3.2 | <i>Task Performance</i> | 18 |
| 2.3.3 | <i>EEG</i> | 21 |
| 2.4 | DISCUSSION | 22 |
| 2.4.1 | <i>Summary</i> | 22 |
| 2.4.2 | <i>Interpretation of results</i> | 23 |
| 2.4.3 | <i>Conclusions</i> | 26 |
| CHAPTER 3: A MODEL OF DYNAMIC, WITHIN-TRIAL CONFLICT RESOLUTION | | |
| FOR DECISION MAKING | | 27 |
| 3.1 | INTRODUCTION | 28 |
| 3.1.1 | <i>Conflict and cognitive control</i> | 28 |
| 3.1.2 | <i>Within-trial mechanisms</i> | 30 |
| 3.1.3 | <i>Models of cognitive control</i> | 31 |
| 3.1.4 | <i>Evidence accumulation processes</i> | 33 |
| 3.1.5 | <i>Summary and outline</i> | 34 |
| 3.2 | MODEL DEVELOPMENT | 39 |
| 3.2.1 | <i>Competition between accumulators</i> | 42 |
| 3.2.1.1 | Constrained FFI model | 45 |
| 3.2.1.2 | LCA model | 46 |
| 3.2.2 | <i>Drive to attention mechanisms</i> | 46 |

| | | |
|---------|--|------------|
| 3.2.2.1 | Time-based attention..... | 51 |
| 3.2.2.2 | Control-based attention..... | 52 |
| 3.2.2.3 | Time with noise..... | 52 |
| 3.2.3 | <i>Summary of model variants</i> | 53 |
| 3.3 | EXPERIMENTS | 54 |
| 3.3.1 | <i>Experiment 1</i> | 55 |
| 3.3.1.1 | Methods..... | 55 |
| 3.3.1.2 | Results..... | 59 |
| 3.3.1.3 | Discussion | 68 |
| 3.3.2 | <i>Experiment 2</i> | 70 |
| 3.3.2.1 | Methods..... | 71 |
| 3.3.2.2 | Results..... | 75 |
| 3.3.2.3 | Discussion | 79 |
| 3.3.3 | <i>Experiment 3</i> | 81 |
| 3.3.3.1 | Methods..... | 81 |
| 3.3.3.2 | Results..... | 87 |
| 3.3.3.3 | Discussion | 93 |
| 3.4 | GENERAL DISCUSSION | 95 |
| 3.4.1 | <i>Summary</i> | 95 |
| 3.4.2 | <i>Interpretation of results</i> | 97 |
| 3.4.3 | <i>Conclusions</i> | 103 |
| | CHAPTER 4: INDIVIDUAL DIFFERENCES IN ATTENTION ALLOCATION DURING A 2-DIMENSIONAL INHIBITORY CONTROL TASK..... | 104 |

| | | |
|--|---|------------|
| 4.1 | INTRODUCTION..... | 105 |
| 4.2 | MATERIALS AND METHODS | 107 |
| 4.2.1 | <i>Participants</i> | 107 |
| 4.2.2 | <i>Stimuli and Apparatus</i> | 107 |
| 4.2.3 | <i>Procedure</i> | 109 |
| 4.2.4 | <i>Computational Models</i> | 110 |
| 4.2.5 | <i>Model Fitting and Assessment</i> | 114 |
| 4.3 | RESULTS | 116 |
| 4.3.1 | <i>Behavioral Results</i> | 116 |
| 4.3.2 | <i>Model Results</i> | 117 |
| 4.4 | DISCUSSION | 121 |
| CHAPTER 5: QUANTIFYING MECHANISMS OF COGNITION WITH AN EXPERIMENT AND MODELING ECOSYSTEM..... | | 123 |
| 5.1 | INTRODUCTION | 123 |
| 5.1.1 | <i>Task Selection</i> | 126 |
| 5.1.2 | <i>Model Development</i> | 128 |
| 5.2 | METHODS | 129 |
| 5.2.1 | <i>Task</i> | 129 |
| 5.2.2 | <i>Model</i> | 130 |
| 5.3 | RESULTS | 132 |
| 5.3.1 | <i>Behavioral validation</i> | 132 |
| 5.3.2 | <i>Model validation</i> | 134 |
| 5.3.3 | <i>Model-based individual difference analyses</i> | 136 |

| | | |
|-----|--|------------|
| 5.4 | DISCUSSION | 141 |
| | CHAPTER 6: CONCLUSIONS | 143 |
| | REFERENCES..... | 145 |
| | APPENDIX A: SUPPLEMENTARY MATERIALS FOR CHAPTER 2 | 165 |
| | APPENDIX B: SUPPLEMENTARY MATERIALS FOR CHAPTER 3..... | 171 |

List of Tables

| | |
|--|-----|
| Table 3.1: Summary of free parameters across models in Chapter 3 | 54 |
| Table 3.2: Average accuracy and mean RTs (ms) across participants for Chapter 3, Experiment 1 | 60 |
| Table 3.3: Average mean RTs (ms) and error rates (in parentheses) across participants for Chapter 3, Experiment 2 | 76 |
| Table 3.4: Average accuracy and mean RTs (ms) across participants for Chapter 3, Experiment 3 | 87 |
| Table 4.1: Model parameters and priors in Chapter 4 | 114 |
| Table 5.1: Summary of flanker model free parameters in Chapter 5..... | 132 |

List of Figures

- Figure 2.1: Orientation of implanted leads. Postoperative anterior-posterior (A) and lateral (B) x-ray images. Fusions of coronal (C) and left hemisphere sagittal (D) pre-op T1 anatomical MRI (3T) and 1-month post-op CT images showing lead placement. 12
- Figure 2.2: Weight progression. Starting point is 335 lbs. at the pre-surgical baseline. Points correspond to individual weight measurements. Line colors correspond to the long-term device settings that were active in the period of time prior to each weight measurement. The horizontal line at 305 lbs. represents 15% excess body weight loss, the *a priori* threshold for determining responder status. OTHER: Points 1–6 include pre-surgical baseline and post-surgery recovery when no stimulation occurred; stimulation parameters could not be verified between points 14 and 15 and between points 20 and 21; insufficient titration and long-term data for evaluating active stimulation parameters between points 9 and 11..... 17
- Figure 2.3: Effects of stimulation on weight loss and flanker performance. A) Rate of long term weight loss (mean lbs. per day) for each set of stimulation parameters. B) Flanker task performance within-session. The horizontal axis is the log-transformed trial number, which is an indicator of how much time had passed since the relevant stimulation parameter set became active. The vertical axis is the patient’s log-transformed reaction time on each trial, which was our dependent variable metric for the patient’s performance on the flanker task. 20

Figure 2.4: t-statistics from the EEG analysis. We performed an ANOVA at each electrode to predict EEG voltage. Log trial number and DBS status (ON, OFF) were factors. The interaction of the factors was a significant predictor of EEG voltage in the left and right frontal quadrants of the participant's scalp. Swaths of color represent t-values from the ANOVA within 5 sub-windows of time after the stimulus appeared..... 22

Figure 3.1: Flowchart of alternative model mechanisms. Each of the four models in our main investigation contained a different combination of mechanisms for attentional focus (time-based vs. cognitive control-based, Panel 1) and evidence accumulation (strongly-correlated vs. weakly correlated, Panel 4). Across all models, an attentional spotlight represented as a density function for a Gaussian distribution (Panel 2) shrinks throughout a trial. Drift rates are calculated from the area under the spotlight allocated to the target and flankers (Panel 3). Evidence is calculated within either an FFI or LCA framework (Panel 4). For control-based models, cognitive control is calculated as the cumulative distance between total evidence and a threshold (Panel 5). This measure is in turn used to calculate the standard deviation of the attentional spotlight in the control-based models, whereas the spotlight shrinks at a constant rate in the time-based models..... 36

Figure 3.2: Diagram of the dual-stage two-phase (DSTP) model. In Stage 1 (left panel), the stimulus selection and response selection phases are represented by racing diffusion processes. If the response selection phase finishes first, a response is made based only on the dominant perceptual features in the stimulus array. If the stimulus selection phase finishes first, no response is made in Stage 1, and either the target or the flankers are selected for controlled attentional processing. In Stage 2 (right panel), the response selection phase drift rate changes to reflect the outcome of the stimulus selection phase..... 41

Figure 3.3: Comparison of FFI and LCA mechanisms. Left column: Graphical models of FFI (A) and LCA (D) processes from stimulus input to response, where dashed lines represent loss of evidence, open circles represent inhibition. Middle column: Simulated paths of evidence accumulation in FFI (B) and LCA (E) for two options in a single trial of a two-alternative choice task. Right column: Phase plane plots of the same decision illustrated in panels B and E for the constrained FFI (C) and the LCA model (F). Black lines show the path of the decision process in a single trial by plotting evidence for each option against one another where 1.0 on each axis represents the decision threshold..... 44

Figure 3.4: Control- and time-based signals to attention. Across models in our comparison, attentional spotlights shrink as a function of control (Left panel) or time (Right panel). Mean simulations of control and time signals are shown for a single trial in the congruent and incongruent conditions..... 49

Figure 3.5: Model-generated spotlight widths through time. For each model, 50 trials were simulated from one participant's best-fitting parameters. Panels show calculations of spotlight standard deviations through time, with each simulation displayed as a gray line to demonstrate the between-trial variability captured by each model. A single additional simulation is shown as a red line to illustrate differences in within-trial variability..... 50

Figure 3.6: Heat map of BPIC values, mean-centered within-subject for Experiment 1. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right. 61

Figure 3.7: Observed and model-generated choice-RT distributions. Observed RT distributions for correct (light gray histograms) and incorrect (dark gray histograms) responses were

averaged across participants. Models were simulated 10,000 times for each condition, using each participant’s best-fitting parameters. Lines show average model-generated distributions across participants. Distributions generated by the FFI time and FFI control models are shown in the top row, whereas distributions generated by the LCA time and LCA control models are shown in the bottom row..... 62

Figure 3.8: Observed and predicted ELI values for incongruent trials. ELI values calculated from each subject’s data in the incongruent condition (x-axis) are plotted against the ELI values generated from each subject’s best-fitting parameters (y-axis) in each model (panels). Correlations and lines of best fit are displayed on each panel..... 64

Figure 3.9: Observed and predicted CAFs for incongruent trials. Data from each subject were sorted according to RT within 6 equally-spaced percentile bins. Performance and minimum RT for each bin were averaged across participants (red Xs). After generating 1,000 choice-RT pairs from each subject’s best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines)..... 65

Figure 3.10: Observed and predicted ELI values for congruent trials. ELI values calculated from each subject’s data in the congruent condition (x-axis) are plotted against the ELI values generated from each subject’s best-fitting parameters (y-axis) in each model (panels). Correlations and lines of best fit are displayed on each panel..... 66

Figure 3.11: Observed and predicted CAFs for congruent trials across low-ELI participants. Data from each subject were sorted according to RT within 6 equally-spaced percentile bins. Performance and minimum RT for each bin were averaged across participants (blue Xs). After generating 1,000 choice-RT pairs from each subject’s best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines).. 67

Figure 3.12: Observed and predicted CAFs for congruent trials across high-ELI participants.

Data from each subject were sorted according to RT within 6 equally-spaced percentile bins. Performance and minimum RT for each bin were averaged across participants (blue Xs). After generating 1,000 choice-RT pairs from each subject's best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines)..... 68

Figure 3.13: Examples of stimuli used in Experiment 2, based on Figure 2 in Servant et al., 2014.

Each stimulus consisted of a target circle (red or blue), flanked by two circles of an incongruent (Left column) or congruent (Right column) color. Targets varied in saturation between 15 and 80% (rows) while the color saturation of the flankers was held constant at 80%. While only stimuli with red targets are shown here, the paradigm was counterbalanced so that 50% of stimuli featured a blue target. 72

Figure 3.14: Range of sigmoid functions for calculating p_c . Sigmoid functions were implemented

to capture the attention allocated to stimuli in the six saturation conditions in Experiment 2. Panel A shows the effect of modifying the a parameter while keeping b and c constant. Panels B and C similarly show the effects of modifying the b and c parameters respectively, while the other parameters are held constant. 74

Figure 3.15: Heat map of BPIC values, mean-centered within-subject for Chapter 3, Experiment

2. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right. 77

Figure 3.16: Observed and model-generated choice-RT distributions. Observed RT distributions

for correct (light gray histograms) and incorrect (dark gray histograms) responses were

averaged across participants. Models were simulated 10,000 times for each condition, using each participant’s best-fitting parameters. Lines show average model-generated distributions across participants. Distributions generated by the FFI time and FFI control models are shown in the left panel, and distributions generated by the LCA time and LCA control models are shown in the right panel. Choice-RT distributions for low target saturation trials are shown in the top row and high saturation trials are shown in the bottom row. 79

Figure 3.17: Condition-level EEG results for Experiment 3. Topographic maps show voltage differences between congruent and incongruent conditions at 370 ms post-stimulus, before (Panel A) and after (Panel B) threshold-free cluster enhancement (TFCE). Panel C shows grand average ERP stimulus-locked waveforms for congruent and incongruent trials at electrode CPz. Significant condition-level differences as determined by TFCE are shown as green points. 88

Figure 3.18: Observed EEG voltages and model-generated attention modulation signals. Data and simulations are shown for one subject. Analyses were completed at the level of every trial and electrode, but for the purposes of this visualization, EEG voltages were averaged across electrodes that demonstrated the highest correlation with model-generated attention signals (TP8, P2, C6, CP6, CPz, Pz, FC6, C2, CP1, T8, P1, P4, FC4). Data were divided into three bins based on three equal RT percentiles. Vertical lines represent the boundaries of the decision-relevant interval between stimulus onset and the mean RT within-bin, limited by the mean best-fitting t_0 across models. 90

Figure 3.19: Mean Z correlation maps for observed EEG data and model-generated attention modulation signals. Data were generated by each model using each participant’s best-fitting

parameters. For each trial, we calculated an average vector of drive to the attention mechanism through time using each model’s simulations. Trial-level correlations between EEG voltage and model-generated attention were calculated. Pearson’s r values were Fisher’s Z-transformed, and p values were calculated for each model and electrode using a 1-sample t-test. Significance was determined via Benjamini-Hochberg correction for multiple comparisons, and are indicated by yellow points. 91

Figure 3.20: Mean Z correlation difference maps for observed EEG data and model-generated attention modulation signals. After calculating Z correlation values for each model and each electrode, we calculated the pairwise difference topographic maps for each possible pair of models. P values were calculated for each model comparison and electrode using a 1-sample t-test. Significant correlation differences were identified using a Benjamini-Hochberg correction for multiple comparisons, indicated by yellow points..... 93

Figure 3.21: Rank order sums of BPIC values for each model and experiment. The best-fitting model for each subject and experiment as determined by BPIC was assigned a rank of ‘1’, the second best model was ranked ‘2,’ and so on. Rank values were summed within-experiment and normed based on the number of subjects in each experiment. Black points indicate mean normed rank order sums across experiments..... 97

Figure 4.1: Examples of stimuli. Colors are used to highlight the contrasting directions configurations of left (red) and right (blue) arrows. All stimuli shown here contain a left-facing target, but analogous stimuli with right-facing targets were included in the experiment as well. 108

Figure 4.2: Representation of the shrinking 2-dimensional spotlight of visual attention. Over the course of a trial, illustrated from left to right, the spotlight shrinks and focuses on the center

arrow of a stimulus array, rendered in each subplot at $z=0$. The strength of visual attention allocated to each arrow in a stimulus array is calculated from the density of a bivariate normal distribution within the corresponding unit square. The top panels show a circular spotlight, such that the shape parameters of the bivariate normal are constrained to be equal. The bottom panels show an elliptical spotlight, such that the two shape parameters are free to vary. 112

Figure 4.3: Behavioral results. Left Panel: Observed and model-predicted scores within condition. Bars show observed mean scores. Triangular and circular points show mean scores predicted by the elliptical and circular spotlight models, respectively. Error bars show 95% confidence intervals of scores across subjects. Right Panel: Score difference, horizontal vs. vertical conditions. Differences in task scores in the horizontal and vertical task conditions, calculated within-subject. 117

Figure 4.4: Subject-level differences in BIC values between the circular spotlight and elliptical spotlight models. The elliptical model outperforms the circular model for the majority of subjects (lower BIC values indicate better fit). 118

Figure 4.5: Model Results. Left Panel: Subject-level scaled differences in best-fitting horizontal and vertical shape parameters for the spotlight in the elliptical spotlight model. Right Panel: Behavior vs. parameter estimates. Spotlight shape asymmetry predicts behavioral performance in the horizontal and vertical conditions. 119

Figure 5.1: Illustration of stimuli in each condition of the flanker task (i.e. easy, moderate, and hard). A) Possible stimulus configurations. Colors were used to highlight the contrasting orientations of the arrows, but stimuli in the actual task were presented in white font against

a dark gray background. B) Possible stimulus locations. Black asterisks show the 8 locations in which a stimulus could have appeared on each trial. 130

Figure 5.2: Performance differences between a standard flanker paradigm and the SUPREME variant. Bar plots show means and 95% confidence intervals of accuracy and performance across 26 subjects who completed 192 trials of a standard flanker paradigm (A) compared to the flanker task variant included in SUPREME (B). 133

Figure 5.3: Posterior distributions for LCA-control model parameters in Experiments 1 from Chapter 3 and the SUPREME study in Chapter 5. The same model was fit to data from 26 subjects in the standard flanker task and the SUPREME flanker task variant. Parameter values were mean-centered within-subject, and distributions were averaged across subjects. 135

Figure 5.4: SUPREME flanker model results. A) Observed and model-generated scores in each condition of the flanker task. Models were simulated 10,000 times for each condition using each participant’s best-fitting parameters. Scores were calculated based on speed and accuracy. Observed mean and 95% CIs of scores are shown as colored bars. Mean and 95% CIs for model-generated scores are shown as gray bars. B) Correlation between observed and model-generated scores across conditions. 136

Figure 5.5: Observed scores and PPDs of low-performing (A and B) and high-performing (C and D) participants. Observed scores are displays as black points, and PPDs are shown as violin plots. Meaningful differences were identified based on percent overlap of PPDs. 138

Figure 5.6: Comparison of posterior parameter distributions for two low-performing participants.

Facet titles indicate the parameter and the percent overlap between the relevant posterior distributions for participants A and B. 140

Figure 5.7: Comparison of posterior parameter distributions for two high-performing

participants. Facet titles indicate the parameter and the percent overlap between the relevant posterior distributions for participants C and D. 141

Chapter 1: Introduction

1.1 Motivation

Human behaviors are infinitely complex, resulting from multiple levels of biological processes and interactions with the environment. Despite major developments in neuroscience methods for measuring and modulating brain activity over the past decade, there remains an explanatory gap between physiological changes at the level of the neuron and the behaviors that manifest downstream (Montague, Dolan, Friston, & Dayan, 2012). This incomplete knowledge about the relationship between brain and behavior is particularly problematic in the clinical setting, where understanding the root causes of an illness or injury is critical for identifying effective treatments. While drug interventions operate at the level of neurotransmitters, clinical diagnoses and treatments are ultimately based on behaviors and subjective reports (Wiecki, Poland, & Frank, 2015). To address these issues, the burgeoning field of computational psychiatry aims to mathematically define how the brain solves specific problems, and in the process, understand how mechanisms could be disrupted by illness or injury (Adams, Huys, & Roiser, 2016). Computational psychiatry encompasses data-driven approaches like machine-learning for identifying discriminable features between groups, and theory-driven approaches like generative modeling for mathematically specifying the neural computations that precede overt behaviors (Maia, Huys, & Frank, 2017; Yao, et al., 2018). Using theory-driven computational methods, the

goal of my dissertation work is to understand and quantify the biologically relevant mechanisms underlying behavior, with a focus on cognitive control.

Cognitive control is a necessary set of functions in tasks involving planning, error detection, novelty, difficulty, and conflict: situations where relying on habitual behaviors are insufficient for optimal performance (Norman & Shallice, 1986). While it is known that cognitive control fluctuates as we complete the tasks of the day and upregulates attention as we encounter competing sources of information, the mechanisms and time courses of these processes remain a topic of active research. In the lab, questions about how and when cognitive control is mobilized are often investigated using speeded reaction time (RT) tasks that require inhibition of an automatic response. A well-studied example comes from the flanker task (Eriksen & Eriksen, 1974; Kopp, Rist, & Mattler, 1996), in which participants are asked to indicate the direction of a central arrow while ignoring distractors that may be incongruent (<<<◇<<<) or congruent (>>>>>>) to the target. As a result of conflicting evidence for the correct response, participants exhibit *congruency effects* of slower and less accurate responses to incongruent trials compared to congruent (Gratton, Coles, & Donchin, 1992).

In the interest of quantifying mechanisms of cognitive control, the flanker task serves as a centerpiece of the presented work. As illustrated by robust behavioral and neuroimaging effects across decades of research, the flanker task taps into the crossroads between low-level perceptual processing and high-level cognitive control. This, in combination with the objective, speeded format of the general paradigm made the flanker task an ideal testbed for model-based investigations of cognitive control and subsequent development of tools for subject-level

mechanistic assessment. My dissertation uses performance on variants of the flanker task to: 1) characterize the relationship between neural circuitry and global cognitive control; 2) develop and compare models representing alternative theories about the role of cognitive control in decision making using behavioral and electroencephalography (EEG) data; 3) investigate individual differences in the spatial allocation of visual attention under conditions requiring cognitive control and 4) present methods for quantifying and tracking latent cognitive control mechanisms through time, both within and between subjects.

1.2 Organization

In Chapter 2, I will present work that was done in collaboration with Per Sederberg and the Center for Neuromodulation at The Ohio State University, and was recently published in *Frontiers in Psychiatry* (Weichart, Sederberg, Sammartino, Krishna, Corrigan, & Rezai, 2020). As a precursor to the modeling work in Chapters 3-5, the goal of Chapter 2 is to demonstrate that the flanker task can access the brain's capacity for cognitive control on a moment-to-moment basis. Previous studies have validated the flanker task for clinical use by confirming expected group-level differences in task-related behaviors requiring cognitive control. For example, adolescents with attention deficit hyperactivity disorder, who have difficulty focusing attention and ignoring distractions in their daily lives compared to healthy controls, also show reduced inhibitory control during incongruent flanker trials as measured by EEG (Wild-Wall, Oades, Schmidt-Wessels, Christiansen, & Falkenstein, 2009). My collaborators and I had the opportunity to test the validity of the flanker task in a novel way, via a longitudinal deep brain stimulation (DBS) case study. In our study, one patient with morbid, treatment-refractory obesity received DBS of the bilateral nucleus accumbens (NAcc) for the modulation of cognitive control

networks. I administered the flanker task while stimulation settings were sampled during device programming (i.e. active contacts, amplitude), then related acute changes in task performance to long-term treatment effects. The results show that flanker task performance indeed tracks acute changes in cognitive control as a direct result stimulation, and maps onto global treatment-related changes in the long term.

Chapter 3 features a set of model comparison studies that was published in *Psychological Review* (Weichart, Turner, & Sederberg, 2020). Given that the flanker task appeared to tap into acute as well as global aspects of cognitive control in Chapter 2, the goal of Chapter 3 was to gain a deeper understanding of the mechanisms underlying behavior on the flanker task. The *spotlight* view of attention was the theoretical core of the project, which suggests that spatial attention is a finite resource that narrows in scope and increases in acuity during visual processing (Eriksen & St. James, 1986). With guidance from Per Sederberg and Brandon Turner, I systematically modified an existing model instantiation of the spotlight view (White, Ratcliff, & Starns, 2011) to investigate competing hypotheses for how decisions are made under conditions of perceptual conflict. Two dimensions of mechanisms were of particular interest: 1) the force that controls within-trial modulation of the attentional system and 2) the mode of competition between different sources of information during the task.

In Chapter 4, I present a manuscript that is under review *Attention, Perception, & Psychophysics* as part of a special issue to honor Charles W. Eriksen. Through the work presented in Chapter 3, I identified a set of mechanisms that translated Eriksen's verbal theory of the attentional spotlight into a generative model that successfully fit data across three experiments. In all three

experiments in Chapter 3, however, stimuli consisted of a single row of items, such that the spotlight was parameterized to handle 1-dimensional spatial inputs only. This is a departure from the verbal theory that describes a 2-dimensional spotlight, and there are several lines of behavioral evidence suggesting that the spotlight naturally takes on the shape of a horizontally-biased ellipse (Andersen & Kramer, 1993; Feng, Jiang, & He, 2007; Hüttermann, Memmert, & Simons, 2014). Working with Per Sederberg, I reconfigured the best-fitting model from Chapter 3 to operate in 2-dimensions and developed a variant of the flanker task in which stimuli were 2-dimensional arrays. Model comparison methods uniquely allowed us to investigate individual differences in the shape of the attentional spotlight, which were potentially masked in the group-level analyses performed in previous work.

Chapter 5 focuses on packaging my work from Chapters 3 and 4 into tools for quantifying cognitive control, and includes portions of a manuscript that was submitted to Behavior Research Methods. Relating back to computational psychiatry, the broader goal of my dissertation work is to develop a means for quickly and objectively evaluating aspects of cognition in terms of biologically relevant mechanisms rather than behavior alone. In cognitive assessments of clinical populations, it is common to interpret performance scores relative to *a priori* cut-offs that are assumed to indicate impairment (Holdnack, et al., 2017). This method of assessment may result in gross over- or underestimation of within-participant changes in performance across time, depending on the relative distance to a diagnostic boundary. We need better ways of accounting for uncertainty in our measurements when interpreting cognitive measures, rather than working exclusively with single-value performance scores and normative standards. To this end, Per Sederberg and his lab have created a cognitive assessment toolbox called SUPREME (Sensing to

Understanding and Prediction Realized via an Experiment and Modeling Ecosystem). My contribution to the project was the 2-dimensional flanker task presented in Chapter 4 and the associated model presented in Chapter 3, as well as validation analyses for all 4 task-model pairs included in SUPREME. Chapter 5 provides the details and results of the flanker task portion of a 2-session validation study, as well as model-based methods for identifying meaningful cognitive changes within and between individuals.

Chapter 2: Cognitive Task Performance during Titration Predicts Deep Brain Stimulation Treatment Efficacy

Published in *Frontiers in Psychiatry*, February 2020

As it pertains to my dissertation work, the purpose of this study was to determine if the flanker task is a viable way of accessing cognitive control at the level of an individual participant. We measured one female participant's ability to engage cognitive control via the flanker task while she independently underwent standard device titration procedures for DBS of the NAcc for the treatment of obesity. Our primary hypothesis was that optimal device settings for long-term clinical outcomes would be retrospectively linked to acute improvement in flanker task performance during titration. Confirmation of our hypothesis demonstrates the sensitivity of the flanker task for within-subject assessments, and provides support for the modeling endeavors presented in Chapters 3-5.

2.1 Introduction

In the context of DBS, 'titration' is the process of adjusting stimulation parameters to reduce symptoms and avoid side effects. For movement disorders, DBS titration is typically

accomplished through trial-and-error methods whereby clinicians sample combinations of device settings (i.e. electrode polarity, amplitude, pulse width, and frequency) and assesses acute clinical effects (Volkman et al., 2002, 2006). Trial-and-error methods have been successful when there is immediate, observable feedback (e.g. alleviation of Parkinson's tremor) following device adjustment. When treating disorders of behavioral rather than movement regulation, however, trial-and-error methods become problematic. In contrast to the physical symptoms associated with movement disorders, behavioral disorders often do not include symptoms that can be objectively observed and measured in the clinical setting. The effects of stimulation can often take weeks or months to manifest (McIntyre et al., 2004), and it can take up to 1-2 years to determine the therapeutic window for stimulation settings (e.g. Figeo et al., 2013).

There is an urgent need for a method of DBS titration that 1) relies on immediate effects with a latency of few minutes rather than weeks or months, 2) is objective, valid, and reliable, 3) is sensitive to incremental stimulation adjustments, 4) can be administered multiple times within a session without response biases, and importantly 5) predicts long-term clinical results. Here, we propose a cognitive task-based method for acute stimulation assessment during NAcc DBS titration. In light of compelling evidence that cognitive performance is sensitive to stimulation loci and strengths (Ezzyat et al., 2017), we investigated the possibility of using objective cognitive measures to guide the selection of optimal stimulation settings for NAcc DBS. Specifically, we chose an inhibitory control task to capture cognitive changes associated with different sets of stimulation parameters. Inhibitory control is broadly defined as the ability to suppress information that interferes with goal-driven behavior (Baumeister et al., 2014). Several lines of evidence have demonstrated that the NAcc plays a critical role in the complex

mechanisms underlying inhibitory control, including lesion studies in rats and local field potential studies in humans (Floresco, 2015).

In the current study, we measured one female participant's ability to engage inhibitory control via the flanker task (Eriksen & Eriksen, 1974) while she independently underwent standard device titration procedures for DBS of the bilateral NAcc. We selected the flanker task for the current project in light of behavioral and electrophysiological evidence that obese participants have slower reaction times and reduced error-related negativity EEG activity during inhibitory control tasks compared to healthy controls (flanker task: Kamijo et al., 2014; Stroop task: Gunstad et al., 2007, Smith et al., 2010; Go- No Go task: Mobbs et al., 2011). Additionally, functional magnetic resonance imaging (fMRI) work has shown that obese participants have reduced inhibitory control activity in the dorsolateral prefrontal cortex (dlPFC; Le et al., 2006) and anterior cingulate cortex (ACC; Volkow et al., 2008). These two regions have been exhaustively studied using the flanker task, with results demonstrating a direct relationship between dlPFC-ACC coactivation and flanker task performance (Luks et al., 2010; Huyser et al., 2011; Fassbender, 2006; Clayson & Larson, 2011).

Data collected during titration was analyzed retrospectively, after the patient had been identified as a DBS responder. Our primary hypothesis was that optimal device settings for long-term clinical outcomes would be linked to acute improvement in task performance during titration. This would support the idea that cognitive testing is a viable alternative to traditional methods of device titration, and is a worthwhile avenue for investigation in future work with a larger cohort of patients. After identifying clinically effective stimulation settings, we collected EEG data

from the patient to gain additional insight into the mechanisms underlying the observed effects of stimulation. Given that DBS of the NAcc has been successfully implemented as a treatment for other behavioral disorders by regulating the frontal-thalamic pathway (Passamonti et al., 2009; Sturm et al., 2003), we hypothesized that optimal device settings would result in increased frontal engagement during the flanker task as measured by EEG. This study serves as a first step toward developing a task-guided tool for DBS titration that has the potential to drastically improve the quality and efficiency of standard procedures for treating behavioral disorders.

2.2 Materials and Methods

2.2.1 Participants

One female participant completed this non-randomized phase I safety and feasibility prospective open label interventional pilot investigating DBS as a treatment for obesity. Two other participants enrolled, but did not complete the study (Rezai et al., 2018) and did not reach responder criteria prior to the time of withdrawal. All participants met or exceeded the 40 kg/m² body mass index classification threshold for morbid obesity, were at least 24 months post Roux-en-Y gastric bypass surgery without evidence of sustained weight loss, and were free of neurological or other severe medical conditions. Participants were recruited upon referral from a nutritionist. Magnetic resonance imaging (MRI) scans at baseline confirmed that participants had no damage to the NAcc target in either hemisphere. The study met institutional requirements for research involving human subjects, was approved by the Food and Drug Administration and Ohio State University's Biomedical Sciences Human Subject Institutional Review Board (Protocol: 2011H0329), and was registered on ClinicalTrials.gov (Identifier: NCT01512134). All participants provided written informed after a full explanation of study procedures by a clinician.

The FDA approved an *a priori* responder criterion for minimal clinical effectiveness of 15% excess bodyweight loss. Details provided in the sections to follow refer only to the participant who completed the full study.

2.2.2 Surgery

DBS leads (Medtronic Neurological Model 3391) were implanted bilaterally in the NAcc via frame-based stereotactic procedures. The anatomical target was visualized using standard axial, coronal, and sagittal T1, T2 and inversion-recovery MRI-guided methods. The Medtronic Stealth navigation system (Stealth Framelink software; Medtronic Inc., Minneapolis, MN) was used to simulate the planned lead trajectory and confirm avoidance of vasculature. Ventral contact locations relative to the midcommisural point were: left hemisphere X=-6.51mm, Y=15.51mm, Z=-5.22mm; right hemisphere X=7.36mm, Y=13.60mm, Z=-5.36mm, where X is medial-lateral, Y is anterior-posterior, and Z is rostral-caudal. Lead locations within the NAcc were verified during surgery using single-cell microelectrode recordings, and were subsequently confirmed using a fusion of pre-op T1 anatomical MRI (3T) and 1-month post-op computed tomography (CT) scans. Anatomical reconstructions of lead placements were confirmed by an expert neuroanatomist. Reconstructed CT/MRI images of the implanted leads are shown in Figure 2.1.

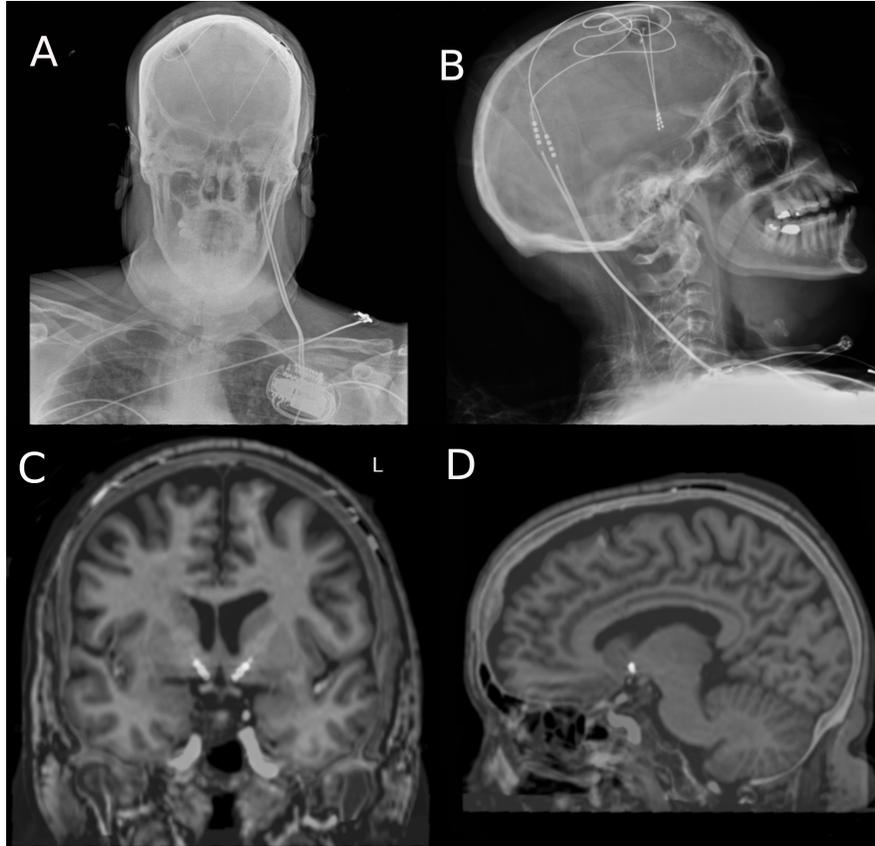


Figure 2.1: Orientation of implanted leads. Postoperative anterior-posterior (A) and lateral (B) x-ray images. Fusions of coronal (C) and left hemisphere sagittal (D) pre-op T1 anatomical MRI (3T) and 1-month post-op CT images showing lead placement.

2.2.3 Titration

Following a 6-week post-surgery recovery phase, the participant attended weekly 1- to 2-hour study visits with a physician for 12 weeks. During these study visits, device settings including polarity, amplitude, pulse width, and frequency were titrated based on the participant's ratings of mood, energy, and anxiety, and avoidance of adverse side effects, per standard procedures. The physician used a systematic, iterative method to sample the parameter space of device settings during each visit. Contact stimulation could be monopolar or bipolar, and up to two contacts

were stimulated in each hemisphere at once. Exploratory settings were applied for 5-15 minutes each.

During this procedure, the participant was intermittently asked to complete 1-5 blocks of the flanker task. The participant indicated the direction of a central arrow while ignoring congruent, incongruent, or neutral (ooo<ooo) distractors. For the incongruent condition specifically, participants in the flanker task need to engage inhibitory control mechanisms in order to overcome influence from the distractor arrows and correctly identify the direction of the target. Each block contained 36 trials, 12 from each condition. A custom program using the State Machine Interface Library for Experiments (SMILE; <https://github.com/compmem/smile>) generated the randomized task lists, presented stimuli, and logged responses. Stimuli were presented on a standard 14-inch laptop screen. The participant pressed the 'J' and 'K' keys on the keyboard to indicate left and right target directions respectively. Stimuli remained on the screen until a response was made, and a fixation cross appeared for a jittered duration of 0.5-1.0s in between stimuli. Cognitive testing via the flanker task was completely independent of device titration, such that the physician was not able to use the participant's task performance to gauge the effectiveness of any particular device settings. Instead, the clinical research team made decisions about changing device settings based on week-to-week weight changes and the participant's self-reports of mood and behavior.

2.2.4 Long-Term Monitoring

After the titration phase, stimulation was continuous for the first 10 months of the long-term follow-up phase. In an effort to conserve the battery life of the device, we introduced a brief trial

period of “cycling” stimulation during which the device was turned on during the day and automatically turned off at night. During this trial, we noted substantially-reduced battery consumption without negative reports from the patient, nor adverse effects on weight-loss. For the latter 20 months of the long-term follow-up phase, we therefore set the device to cycle between a 16-17-hour ‘ON’ state during the day and a 7-8-hour ‘OFF’ state at night. The participant attended monthly study visits for weight and body fat percentage measurements, nutritional counseling, and adverse effects monitoring. Stimulation parameters were adjusted as needed, based on participant feedback and in an effort to improve clinical effects. As in the titration phase, the flanker task was administered throughout long-term follow-up study visits if changes to the device settings were made. Weight and active stimulation settings were recorded each time a change was made, or at least once per month when settings were stable. Given that substantially more unique stimulation settings were tested during titration compared to long-term monitoring, stimulation settings were organized into bins based on active contacts and “high” or “low” voltage relative to 5V for the purposes of our analyses. Three sets of stimulation parameters were represented in both the titration phase and the long-term monitoring phase: 1) bilateral lower middle contacts, low amplitudes; 2) bilateral lower middle contacts, high amplitudes; and 3) both bilateral middle contacts, high amplitudes.

2.2.5 EEG Recording

The participant completed one EEG session after long term stimulation settings had been in a clinically effective range for 14 days (treatment settings at the time of the session: LEFT: Case+ 1-, 5 V, 120 μ s, 130 Hz; RIGHT: Case+ 9-, 3.5 V, 120 μ s, 130 Hz). The participant completed two blocks of the flanker task (blocks 1-2) in a DBS-ON state. Bilateral stimulation was turned

off using a Medtronic wireless DBS Patient Programmer, and the participant completed two more blocks of the task in the DBS-OFF state (blocks 3-4). Stimulation was turned on again, and the participant completed two final blocks of the flanker task in the DBS-ON state (blocks 5-6). In order to capture acute neural effects of DBS, no more than two minutes passed from the time of switching the device ON or OFF to when a task block began. Blocks 1 and 2 were excluded from analyses in an effort to ensure that our results would be driven by the effects of stimulation rather than practice effects. Here, “practice effects” refer to incidental improvements in performance as participants acclimate to a task (Wenke et al., 2015; Mayr & Awh, 2009).

Stimuli were presented and responses were recorded via a desktop PC connected to a 24” LCD display. The participant was fitted with an elastic cap embedded with 64 Ag-AgCl scalp electrodes arranged in an extended 10-20 array (BrainProducts GmbH, Munich, Germany), and seated in an electrically-shielded testing room. Electrodes were referenced to Cz. The signal was sampled at a rate of 1000 Hz via a DC-powered actiCHamp amplifier connected to a desktop PC. Electrode impedances were reduced to less than 25 Kohms prior to experimental testing, as recommended by the equipment manufacturer. EEG signal was monitored throughout the session for abnormalities using PyCorder software (BrainProducts GmbH, Munich, Germany) on the acquisition PC.

All EEG preprocessing was completed using custom functions using the software package Python Time Series Analysis (PTSA; <https://github.com/compmem/ptsa>). Data were filtered from 0.25-20 Hz to eliminate low-frequency noise and electrical artifacts from the DBS

generator. Wavelet-enhanced independent component analysis (Castellanos & Makarov, 2006) removed artifacts from eye-blinks and saccades.

2.3 Results

2.3.1 Optimal Stimulation Settings for Weight Loss

Approximately 5 months after the device was turned on, the participant reached *a priori* responder criteria of 15% excess weight loss. Over the course of the 36-month study, the participant lost a total of 98.8 pounds equal to 47% of her excess body weight, dropping from an initial body mass index of 55.8 to 39.3. Three sets of stimulation parameters were tested for long-term effects: (1) Bilateral stimulation of ventral-medial contacts at amplitudes less than or equal to 5V (minimum of 2V) resulted in the fastest rate of weight loss (47.80 lbs. lost over the course of 129 days of stimulation, mean change of -0.37 lbs./day). We therefore determined *post hoc* that these were the optimal settings for weight loss in this participant. (2) Stimulating the same contacts at higher amplitudes (>5V, maximum of 9V) were associated with only modest weight loss (1.61 lbs. lost over the course of 108 days of stimulation, mean change of -0.01 lbs./day). (3) Both bilateral medial contacts were stimulated at high amplitudes (>5V, maximum of 8V) for most of the study's duration, resulting in a substantial net weight loss (47.00 lbs. lost over the course of 675 days), but at a less striking rate than optimal (mean change of -0.07 lbs./day). Figure 2.2 shows the participant's weight progression throughout the study.

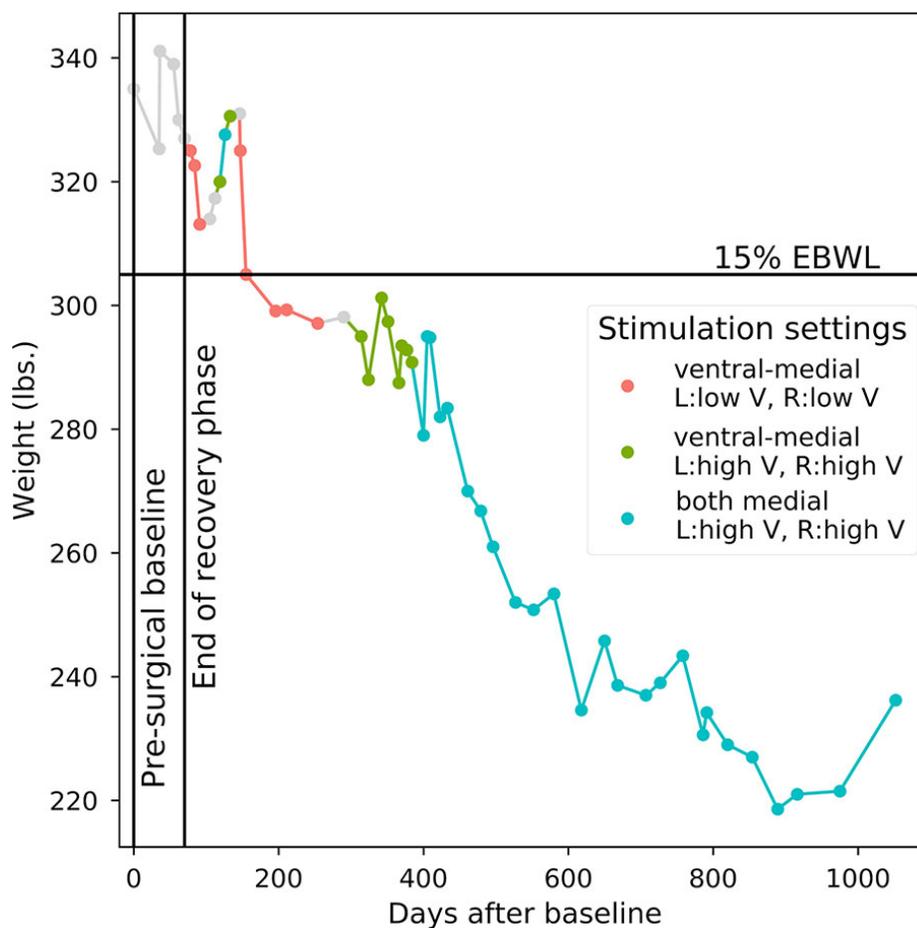


Figure 2.2: Weight progression. Starting point is 335 lbs. at the pre-surgical baseline. Points correspond to individual weight measurements. Line colors correspond to the long-term device settings that were active in the period of time prior to each weight measurement. The horizontal line at 305 lbs. represents 15% excess body weight loss, the *a priori* threshold for determining responder status. OTHER: Points 1–6 include pre-surgical baseline and post-surgery recovery when no stimulation occurred; stimulation parameters could not be verified between points 14 and 15 and between points 20 and 21; insufficient titration and long-term data for evaluating active stimulation parameters between points 9 and 11.

2.3.2 Task Performance

Overall accuracy across all task blocks was near ceiling (accuracy across all conditions: 0.988, accuracy for incongruent trials only: 0.982). Task data from all visits were sorted based on active contacts and stimulation amplitudes relative to 5V. To remove within-session practice effects, the first block from each visit was excluded from further analyses (Wenke et al., 2015; Mayr & Awh, 2009). Because the incongruent task condition is most relevant for measuring inhibitory control, we only considered incongruent trials in our behavioral analyses. Incorrect trials and RT outliers were removed. RTs and within-trial trial numbers were log-transformed in an effort to satisfy normality assumptions (Gelman & Hill, 2006; Zou, Tuncali, & Silverman, 2003). Trial numbers since stimulation change were also log-transformed. Data from each stimulation parameter set were individually compared to DBS-OFF (Xie et al., 2017). Log RTs were analyzed using likelihood ratio tests of mixed-effects models where the factors were DBS status (ON, OFF) and log trial number. By-run intercepts and random slopes for the interaction terms were included as random effects.

Following activation of optimal stimulation parameters (as determined by mean rate of weight loss; bilateral ventral-medial contacts, low amplitudes), the participant made significantly faster correct responses to incongruent task items compared with the device was turned OFF ($X^2(1)=4.571$, $p=0.033$). The participant's RTs were not significantly affected by stimulation with any other parameter sets sampled during titration, including those tested for long-term treatment effects (bilateral ventral-medial contracts, high amplitudes: $X^2(1)=0.301$, $p=0.580$; both bilateral medial contacts, high amplitudes: $X^2(1)=0.255$, $p=0.613$). Long-term rate of weight-loss and acute cognitive performance (as measured by log trial-level RTs in the

incongruent task condition) for each set of stimulation settings are shown in Figure 2.3. Using the same factors and dependent variable, direct comparisons between optimal and sub-optimal parameter sets did not yield statistically significant results, and are reported in Appendix A. Comparing each active stimulation condition to DBS-OFF where the dependent variable was RTs for correct responses in the congruent condition also did not yield statistically significant results (bilateral ventral-medial contacts, low amplitudes: $X^2(1)=1.038$, $p=0.308$; both bilateral medial contacts, high amplitudes: $X^2(1)=2.562$, $p=0.110$; bilateral ventral-medial contacts, high amplitudes: $X^2(1)=0.0045$, $p=0.947$). Because the regression analysis yielded a significant effect for incongruent but not congruent trials while optimal stimulation parameters were active, we believe DBS is specifically affecting inhibitory control rather than response speed in general.

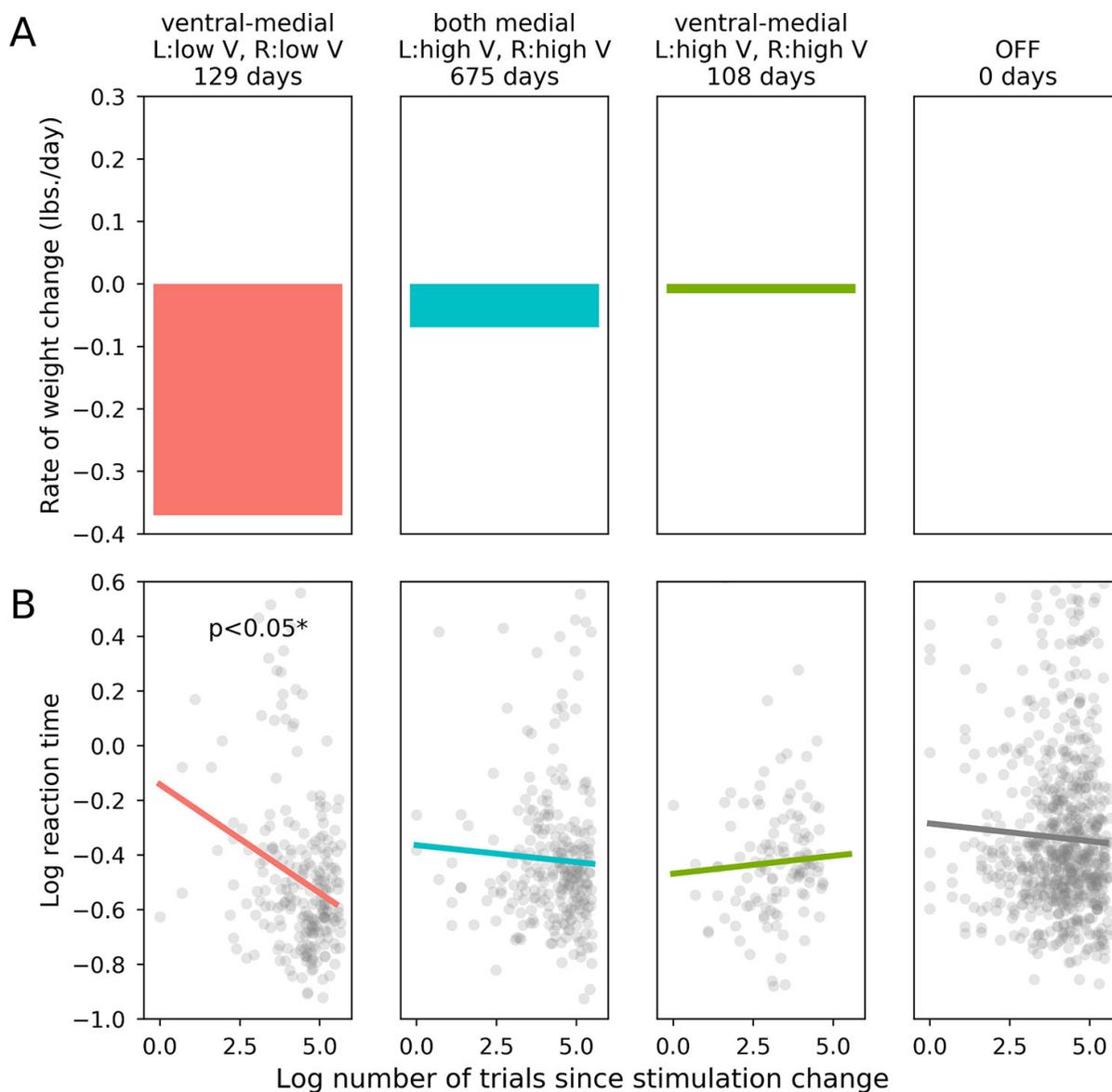


Figure 2.3: Effects of stimulation on weight loss and flanker performance. A) Rate of long term weight loss (mean lbs. per day) for each set of stimulation parameters. B) Flanker task performance within-session. The horizontal axis is the log-transformed trial number, which is an indicator of how much time had passed since the relevant stimulation parameter set became active. The vertical axis is the patient's log-transformed reaction time on each trial, which was our dependent variable metric for the patient's performance on the flanker task.

2.3.3 EEG

Incongruent trials were segmented into events and time-locked to stimulus onset. Events were 3000 ms long, beginning 1500 ms pre-stimulus onset and baseline-adjusted to 100 ms pre-stimulus. Events were rejected if kurtosis exceeded 5.0 or if amplitude range exceeded 100V. Electrodes were grouped into 4 quadrants, representing left anterior, right anterior, left posterior, and right posterior regions respectively (left anterior: F1, F3, FC1, FC3, FC5, C1, C3, C5; right anterior: F2, F4, FC2, FC4, FC6, C2, C4, C6; left posterior: CP1, CP3, CP5, P1, P3, P5, PO3, PO7; right posterior: CP2, CP4, CP6, P2, P4, P6, PO4, PO8).

The participant performed with high accuracy throughout the EEG session (accuracy across all conditions: 0.985, accuracy for incongruent trials only: 0.986). RTs slowed after DBS was turned from ON to OFF (slope=0.067), and became faster when DBS was turned from OFF to ON (slope=-0.021). These results, however, were not statistically significant (ON to OFF: $R^2=0.164$, $p=0.443$, $\alpha=0.05$; OFF to ON: $R^2=-0.037$, $p=0.864$, $\alpha=0.05$). A 2-way ANOVA predicted EEG amplitude from the interaction of DBS status (ON, OFF) and log trial number following a change in stimulation. In order to assess EEG activity during the decision interval within each trial, we defined a post-stimulus time window of interest from 300-400 ms. This window has been selected for assessing voltage differences between congruent and incongruent flanker stimuli in past studies (Yeung et al., 2004), and allowed us to ignore irrelevant artifacts from early perceptual processes and motor-planning. Figure 2.4 shows T-values from the ANOVA at each electrode, split into 5 equal time bins within the window of interest. In both the left and right anterior quadrants, there was a drop in amplitude through time after DBS was turned OFF

compared to when it was turned ON. Correcting for multiple comparisons, these effects were significant (left anterior: $F(28,27)=8.52$, $p=0.007$, $\alpha=0.0125$; right anterior: $F(28,27)=9.23$, $p=0.005$, $\alpha=0.0125$). We did not observe a significant change in amplitude in the left or right posterior quadrants (left posterior: $F(28,27)=4.73$, $p=0.038$, $\alpha=0.0125$; right posterior: $F(28,27)=4.33$, $p=0.047$, $\alpha=0.0125$).

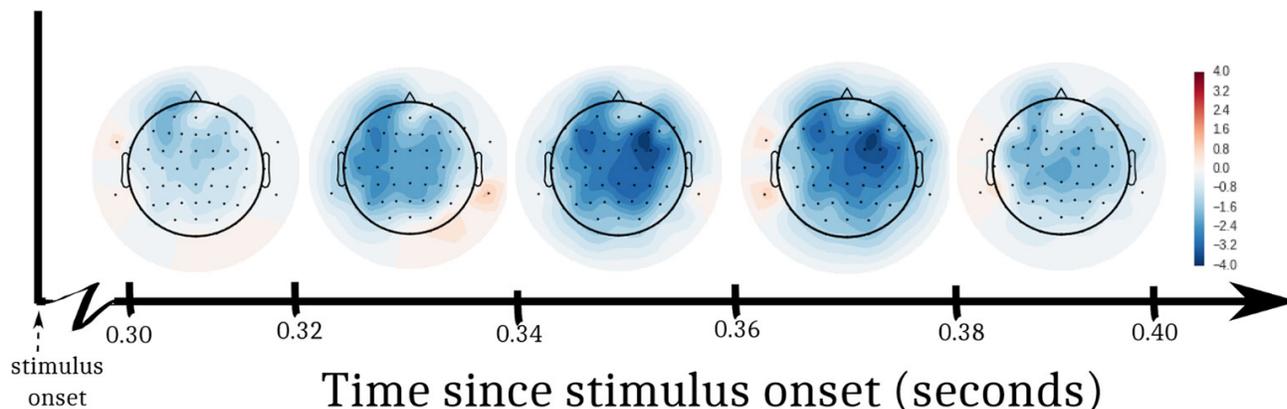


Figure 2.4: t-statistics from the EEG analysis. We performed an ANOVA at each electrode to predict EEG voltage. Log trial number and DBS status (ON, OFF) were factors. The interaction of the factors was a significant predictor of EEG voltage in the left and right frontal quadrants of the participant's scalp. Swaths of color represent t-values from the ANOVA within 5 sub-windows of time after the stimulus appeared.

2.4 Discussion

2.4.1 Summary

Standard trial-and-error methods of DBS device titration depend on immediate, measurable effects of individual sets of stimulation parameters. As clinical applications for DBS have expanded beyond movement disorders, device titration methods have not been adequately

adapted for behavioral disorders lacking overt physical symptoms. While current methods rely on subjective ratings of mood, energy, and anxiety to guide device titration and select parameters for long-term stimulation, we investigated cognitive task performance as a possible alternative. Based on previous work that has defined the role of the NAcc within a complex cognitive architecture (Sesia et al., 2010), we hypothesized that acute performance on an inhibitory control task during device titration could predict long-term treatment efficacy of NAcc DBS. Converging evidence from the current study indeed suggested a link between acute cognitive performance and subsequent clinical outcomes as determined by retrospective analyses.

2.4.2 Interpretation of results

Post-hoc linear mixed effects regression analyses indicated that the DBS settings linked to the fastest rate of weight loss produced an immediate, significant improvement in flanker task performance during device titration. This finding is in line with previous work investigating acute changes in task performance related to different DBS-ON states as a way to tangentially assess stimulation efficacy. Mikos and colleagues (2011), for example, used a computational model-based approach to link volumes of tissue activation at different contacts during DBS of the subthalamic nucleus to letter fluency in Parkinson's disease patients. Their results suggested that cognitive performance correlates with treatment effects in motor disorders. Our results show that this connection potentially holds for behavioral disorders as well, even in cases when treatment results are not immediately observable.

EEG results provided further insight into the neural mechanisms underlying the optimal DBS settings. DBS within the optimal parameter range resulted in a significant difference in cortical

amplitude at frontal electrodes compared to when DBS was OFF. These are the results that we would expect, given that cognitively normal subjects show a higher-amplitude EEG peak at frontocentral electrodes during inhibitory control tasks (Botvinick et al., 1999). We believe our EEG results reflect higher engagement of conflict monitoring processes when optimal DBS settings are active compared to a DBS-OFF state. Further work will need to determine how EEG effects are linked to long-term treatment outcomes, but these results are nevertheless in line with positron emission tomography and fMRI evidence of frontal dysfunctions in obese participants. In particular, obese individuals have reduced activity related to inhibitory control in the dlPFC (Le et al., 2006) and ACC (Volkow et al., 2008). As indicated by our results, DBS of the NAcc may be modulating these frontal networks and thus counteracting this hypoactivity and associated lack of inhibitory control in our participant.

While low amplitude stimulation at ventral-medial contacts was optimal for weight loss, increasing amplitudes above 5V at the same contacts both diminished cognitive task performance and caused the participant's weight loss to slow. Whereas high-amplitude stimulation is often used to achieve treatment effects by mimicking tissue lesions (Blond et al., 1992; Nuttin et al., 1999), this is not necessarily a desirable approach for all cases. Diffusion tensor imaging connectivity analyses, which are included in Appendix A, illustrate why low amplitude stimulation proved to be effective for treatment in this case while high amplitude stimulation did not. Low amplitude stimulation significantly increased connectivity to dorsal attention networks and simultaneously decreased connectivity to the default mode network. High amplitude stimulation, on the other hand, resulted in expansive, nonspecific connectivity without a significant advantage of any network in particular. High-amplitude NAcc DBS has been argued

to benefit obsessive compulsive disorder due to blockade effects within an otherwise hyperactive information processing network connecting the basal ganglia, amygdala, thalamus, and prefrontal cortex (Sturm et al., 2003). For a disorder like obesity that is characterized by a *hypoactive* frontal-thalamic pathway (Passamonti et al., 2009), however, an approach geared toward targeted upregulation rather than attenuation appears to be more appropriate. Weight-loss alongside low-amplitude DBS of the NAcc was also recently observed in a handful of case studies (see Formolo et al., 2019 for recent review; Harat et al., 2016; Mantione et al., 2010; Tronnier et al., 2018).

In order for cognitive testing to be a viable tool for titration, it is important to choose a cognitive task that is relevant to both the stimulation target and the behavioral disorder of interest.

Selecting the flanker task for the present study involved careful consideration of NAcc function and its relationship to obesity. Critically located in the basal forebrain, the NAcc serves as a hub of communication among limbic (ventral tegmental area, substantia nigra, and basolateral amygdala), motor (pallidum and striatum), and executive functioning (prefrontal cortex) networks. Given its proximity to and functional relationships with several key structures in the reward pathway, NAcc stimulation has been proposed to modulate mood, impulsivity, and reward-seeking behaviors via dopaminergic signaling (Goto & Grace, 2005; Grace, 2000). The role of the NAcc in inhibitory control was of particular interest in the present study, with compelling support from animal literature showing that NAcc stimulation affects inhibitory control on an immediate time scale (Sesia et al., 2010). Furthermore, evidence from local field potential recordings in humans showed that inhibitory control paradigms such as the flanker task specifically engage the NAcc (H.-J. Heinze et al., 2009; Voges et al., 2013). Our study aimed to

capitalize on the relationship between the NAcc, inhibitory control, and obesity to link immediate effects of DBS to treatment efficacy.

2.4.3 Conclusions

We propose that task-based titration can be extended beyond the flanker task and the NAcc, and future work will further investigate how we can use acute cognitive performance to predict long-term treatment outcomes. Though the implications of our results are obviously limited due to our sample size, we have provided preliminary evidence that cognitive testing may be a valuable tool for titration.

Chapter 3: A Model of Dynamic, Within-Trial Conflict Resolution for Decision Making

Published in *Psychological Review*, March 2020

For the work in Chapter 2, my colleagues and I were presented with the unique opportunity to directly manipulate attention and control networks via DBS, and measure the patient's subsequent behavior with the flanker task. Given that performance on the flanker task did indeed correlate with global cognitive control as measured by weight loss, the work presented in Chapter 3 aims to gain a deeper understanding of the specific cognitive mechanisms involved in completing the task. Here, I present competing verbal theories for how and when cognitive control is engaged during the task, formalize the theories into mathematical models, and compare theories by fitting the associated models to data from three experiments.

3.1 Introduction

To achieve our goals and navigate a world that is teeming with distractions, humans rely on cognitive control to manipulate limited processing resources in a goal-directed manner. While it is known that cognitive control fluctuates as we complete the tasks of the day and upregulates attention as we encounter competing sources of information, the mechanisms and time courses of these processes remain a topic of active research. In addition to work showing post-feedback modulation of attention via cognitive control to improve future performance (Blais et al., 2007; Botvinick et al., 2004; Verguts & Notebaert, 2008), there is growing evidence that cognitive control acts at faster time scales as well (Braver, 2012; Goschke & Dreisbach, 2008; Ridderinkhof, 2002; Scherbaum et al., 2011). Several mechanisms have been proposed to underlie dynamic changes in attention and cognitive control, including competition between excitatory and inhibitory inputs (Frank, 2006; Scherbaum et al., 2012), asynchrony between processing areas in the brain (Verguts, 2017), and time itself (Hübner et al., 2010; Ulrich et al., 2015; White et al., 2011). Given that all of these mechanisms within their respective computational frameworks can capture aspects of human behavior, substantial overlap in model predictions has made it difficult to draw any stable conclusions about how attentional processes are engaged. In the current study, we investigated the dynamic modulation of attention via cognitive control by developing, fitting, and comparing models representing competing hypotheses for how decisions are made under conditions of perceptual conflict.

3.1.1 Conflict and cognitive control

Growing evidence for moment-to-moment fluctuations in visual attention has led to questions about the impetus and time course of cognitive control. These questions are typically

investigated with paradigms like the flanker task, which require participants to inhibit an automatic response before making a decision. While congruent stimuli only contain evidence for the correct response, incongruent trials require participants to resolve conflict between the target and distractors before making a decision. As a result, participants are slower and less accurate at responding to incongruent trials compared to congruent (Gratton et al., 1992). This *congruency effect* is reduced when incongruent trials occur consecutively, and responses tend to be slower and more accurate following errors. Both results have been interpreted as evidence for modulation of cognitive control as a direct response to the presence of conflict (see Larson et al., 2014 for review).

Influential connectionist modeling work by Botvinick and colleagues (Botvinick, 2007; Botvinick et al., 1999, 2001, 2004; Yeung et al., 2004) suggested that a specialized monitoring center in the brain outputs a measure of conflict at the end of each trial, and subsequently triggers adjustments in cognitive control. After a conflict trial, an increase in cognitive control boosts attentional processing of the goal-relevant target, which in turn improves performance on the next trial. By analyzing flanker task simulation results, the authors found that the output of the conflict monitoring unit in their model resembled typical electroencephalography (EEG) effects, specifically, higher and more sustained peak voltage following errors compared to correct responses (Botvinick et al., 2001). The *conflict monitoring hypothesis* has garnered substantial support from neuroimaging work, localizing conflict detection functions to the ACC and identifying modulation of attentional control within the dlPFC (Kerns et al., 2004; MacDonald et al., 2000; Ridderinkhof et al., 2004; van Veen & Carter, 2002).

3.1.2 Within-trial mechanisms

Other lines of work have questioned the timescale assumed by the conflict monitoring hypothesis. Evidence from behavioral and neurophysiological work has suggested that cognitive control is adjusted *within*-trial, in addition to *after* conflict occurs (Burle et al., 2002; Czernochowski, 2015; Nigbur et al., 2015; Ridderinkhof, 2002). Scherbaum and colleagues (2011), for example, collected EEG data while participants completed a modified flanker task with separate visual frequency tags for targets and distractors. By dissociating the attentional processing signals for the different stimuli, the researchers were able to identify within-trial adjustments in cognitive control alongside the occurrence of conflict, in addition to carry-over cognitive control engagement from previous trials. Alternatives to the conflict monitoring hypothesis have therefore proposed that cognitive control operates on multiple timescales (Braver et al., 2008; J. Brown et al., 2007; Davelaar, 2008). Braver's *dual mechanisms of control* framework (Braver, 2012; DMC; Braver et al., 2008; De Pisapia & Braver, 2006) suggests that cognitive control operates in two modes: a stable 'proactive' mode that biases attention systems to anticipate and prevent conflict, and a variable 'reactive' mode that dynamically detects and resolves conflict as it occurs. Simulations of a DMC connectionist model closely matched behavior and blood oxygenation level dependent (BOLD) imaging data in the ACC and dlPFC during a cognitive control task, and provided evidence of shifting reliance on proactive and reactive control modes between task conditions (De Pisapia & Braver, 2006). As noted by Jiang and colleagues (2014), however, there is still little empirical evidence that the ACC, which has repeatedly been shown to monitor conflict, contains multiple distinct monitoring units operating at different timescales within-trial.

3.1.3 Models of cognitive control

To further delve into within-trial mechanisms independent from carry-over effects from previous trials, theories about cognitive control have also been articulated within the *sequential sampling* class of models (SSMs). Connectionist models are particularly useful for capturing changes over the course of a task such as between-trial congruency effects, due to their complex, interactive architecture and ability to continuously update context (Ratcliff et al., 1999). The flanker SSMs, in contrast, were developed to explain within-trial mechanisms underlying robust *conditional accuracy effects*: faster errors than correct responses in the incongruent condition (Gratton et al., 1988). In general, it is assumed that attention is influenced by distractor items at the beginning of a trial, but focuses on the target as cognitive control is engaged (De Jong et al., 1994; Desimone & Duncan, 1995; Mesulam, 1990). The flanker SSMs offer a range of accounts for how this process unfolds, drawing inspiration from the literature on attention (Hübner et al., 2010; White et al., 2011) and automaticity (Ulrich et al., 2015). Notably, all three of the existing flanker SSMs describe decision and attentional processes that are calculated as a function of time. As such, these models assume cognitive control processes engage based only on the stimulus at hand and the amount of time spent on a trial. This contrasts with the connectionist models, which assume cognitive control is based on layered inputs from continuously-interacting populations of neurons.

In the current article, we introduce an SSM of the flanker task in which cognitive control and attention are emergent properties of the dynamics of the decision itself. Three core concepts from decades of research on cognitive control are foundational to this work: 1) conflict arises from the mutual activation of multiple choice options, 2) cognitive control is deployed as a direct response

to the presence of conflict, and 3) cognitive control biases visual attention toward goal-relevant information. We begin with a standard two-accumulator SSM framework, in which noisy evidence for each possible response accumulates through time until a decision boundary is reached. In our model, a measure of cognitive control is continuously calculated within-trial based on the total amount of evidence across responses. The area of the visually attended region is in turn calculated from the cognitive control output, narrowing onto the target as cognitive control increases or widening as the need for control relaxes away. As in the *shrinking spotlight* (SSP) model introduced by White, Ratcliff, and Starns (2011), the evidence for each response is calculated from the amount of attention allocated to target and distractors, respectively. The proposed model is a closed-loop system, in which cognitive functions are a passive byproduct of interacting processes within the broader decision and action. This framework presents a parsimonious alternative to modularized conflict monitoring and cognitive control in the connectionist models, and also serves as a biologically plausible alternative to the strictly time-based processes in the SSMs.

The idea of cognitive control as an emergent property of activation dynamics has been suggested previously (Mayr & Awh, 2009; Ward & Ward, 2006) and has been implemented in a connectionist model of the flanker task (Scherbaum et al., 2012). The current work stands apart, however, in a number of ways. First, our novel implementation of dynamic processing in an SSM framework allows us to focus on within- rather than between-trial mechanisms. Second, the SSM framework in combination with Bayesian-inspired analysis techniques gives us the power to go beyond generating data that only matches summary statistics, and to fit our model to full distributions choice-RT data at the individual-subject level. This allows us to assess our model's

ability to capture the nuanced differences in performance from subject to subject. Third, we fit multiple model variants representing alternative mechanistic hypotheses to the same sets of observed data, and provide a quantified comparison of goodness-of-fit statistics. Given that nearly all published models are able to match observed data in some capacity, the ability to directly compare fit quality based on full distributions of data is critical for model falsifiability. We did not simply want to determine if a within-trial mechanism for cognitive control could generally capture the data, but rather wanted to identify which specific patterns of subject-level data were better fit by our model compared to a time-based alternative.

3.1.4 Evidence accumulation processes

We developed models with an attentional system driven by cognitive control as will be described in detail in Section 3.2, and compared them to models with an attentional system driven by time as in the SSP developed by White and colleagues (2011). Given that our mechanism of interest critically depends on the evidence for two-choice alternatives, defining the nature of competition between accumulators was a matter of importance. There is considerable discrepancy on this point when comparing the relevant connectionist models to the flanker SSMs. In connectionist models, units representing separate groups of neurons are organized into layers, which in turn correspond to different elements of a trial such as perception, attention, and decision. Units connect to one another in a weighted fashion, passing excitatory or inhibitory inputs from layer to layer. Though units critically affect each other, they typically maintain some level of independence due to random noise, nonlinear activation functions, probabilistic firing, and passive decay of activity (e.g. Liu et al., 2008; McClelland & Cleeremans, 2009). As such, activation of both “left” and “right” decision units in a flanker task may occur simultaneously.

The existing flanker SSMs, however, consider evidence for the two responses to be perfectly anticorrelated, and only evidence for the “left” or the “right” can be above zero at any given time. To compare these assumptions, the models in our investigation included evidence accumulation mechanisms that were either strongly-correlated as in the original flanker SSMs, or were weakly-correlated and governed by leak and lateral inhibition mechanisms to approximate elements of the connectionist framework. Specifically, model variants incorporated calculations from two well-studied SSMs: the *feedforward inhibition* (FFI) model (Shadlen & Newsome, 2001) and the *leaky-competing accumulator* (LCA) model (Usher & McClelland, 2001, 2004).

3.1.5 Summary and outline

In our main comparison, each model contains a combination of mechanisms from two different categories: drive to attentional processes (time-based vs. control-based attentional processing), and competition between accumulators (strongly- vs. weakly- correlated). These alternative mechanisms are illustrated as a flowchart in Figure 3.1. As in the SSP, visual attention is conceptualized as a target-centered density function for a Gaussian distribution. The standard deviation of the attentional spotlight changes throughout a trial, either as a function of time itself or an internal calculation of cognitive control. Drift rates for the two accumulators in the decision process are determined by the area under the attentional spotlight allocated to the target and flankers, respectively. Evidence for each response is calculated within either the FFI or the LCA framework, such that the accumulators are strongly- or weakly- correlated with one another as they stochastically race toward a decision boundary. In the control-based models, cognitive control is represented as the cumulative distance between the total evidence and a threshold, δ . Because the conflict models were designed as a closed-loop system, this measure of cognitive

control feeds back into the calculation of the attentional spotlight standard deviation at the next moment in time.

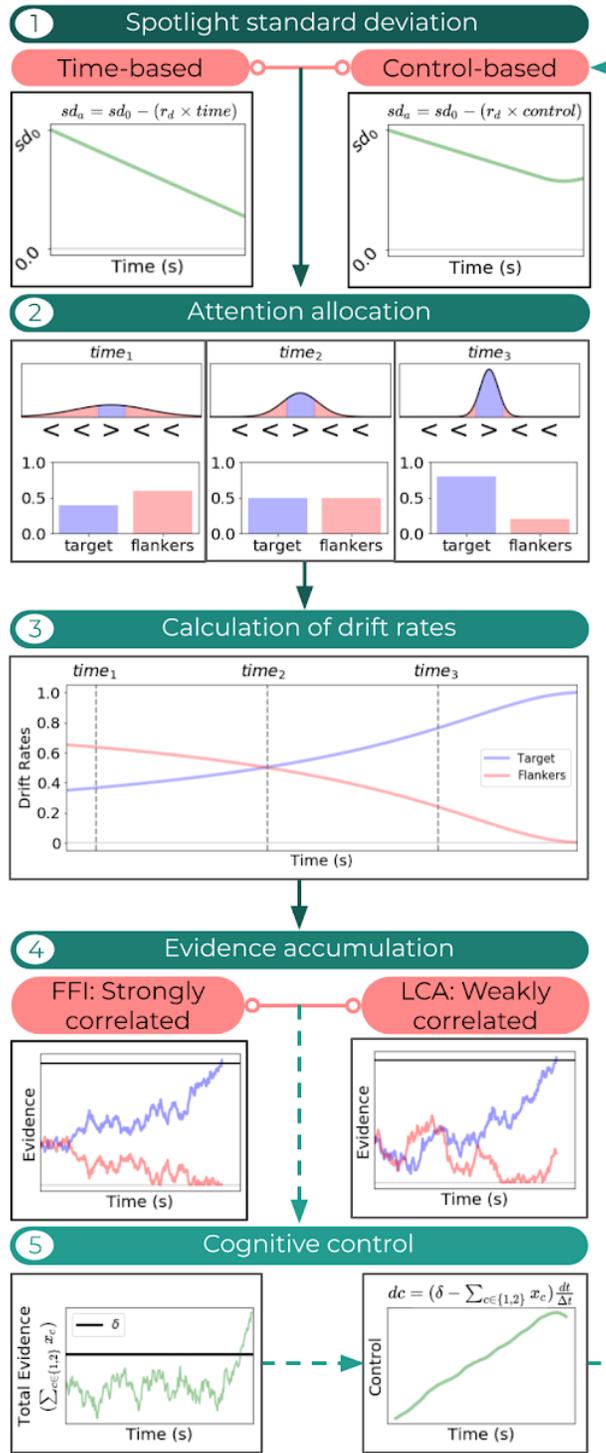


Figure 3.1: Flowchart of alternative model mechanisms. Each of the four models in our main investigation contained a different combination of mechanisms for attentional focus (time-based

vs. cognitive control-based, Panel 1) and evidence accumulation (strongly-correlated vs. weakly correlated, Panel 4). Across all models, an attentional spotlight represented as a density function for a Gaussian distribution (Panel 2) shrinks throughout a trial. Drift rates are calculated from the area under the spotlight allocated to the target and flankers (Panel 3). Evidence is calculated within either an FFI or LCA framework (Panel 4). For control-based models, cognitive control is calculated as the cumulative distance between total evidence and a threshold (Panel 5). This measure is in turn used to calculate the standard deviation of the attentional spotlight in the control-based models, whereas the spotlight shrinks at a constant rate in the time-based models.

We fit all models to data collected in three experiments. Experiment 1 was a standard flanker task with arrow stimuli, in which participants indicated whether a central target was pointing ‘left’ or ‘right’. We were interested in observing how models with dynamic mechanisms for cognitive control would compare to those with time-based mechanisms in the standard paradigm, given that the time-based flanker SSMs have been shown to capture general congruency and conditional accuracy effects in the past (White et al., 2011). In Experiment 2, which was designed and administered by Servant and colleagues (2014), participants were asked to indicate whether a target circle was red or blue while ignoring congruent (same-color) or incongruent (different-color) distractor circles. Importantly, targets varied in color saturation across six different conditions while the color saturation of the flankers was held constant. Here, the models with strongly-correlated accumulation mechanisms would predict equal and opposite evidence for the ‘red’ and ‘blue’ responses across saturation conditions. Models with weakly-correlated accumulation mechanisms governed by leak and lateral inhibition, however, would predict variations in evidence for each response that correspond to the perceptual strength of the relevant

stimulus. In Experiment 3, EEG data were collected as participants completed a standard flanker task. With its high temporal resolution, EEG methods provided insight into the decision process during a standard flanker task that we could not get from behavior alone. Using a model-based EEG analysis with a latent input approach (Mack et al., 2013; Palestro et al., 2018; Turner et al., 2017), we determined the correlation between each model's calculations of attentional drive and observed neural activity at the level of each individual trial. Across these three experiments, we found converging evidence that control-based attention processes in combination with evidence accumulation mechanisms governed by leak and lateral inhibition provided the best fits to behavioral data and uniquely mapped onto observed decision-related signals in the brain.

Our goal was to investigate the possibility of cognitive control as an emergent property of decision dynamics, within a framework that was amenable to data-fitting and quantifiable comparisons. Starting with an existing SSM that was designed to capture the behavioral effects of perceptual conflict, we developed, fit, and compared new model variants that represent competing hypotheses on the nature of within-trial decision processes. We have organized the current article as follows. First, we will provide an overview of the existing SSMs of behavior under conditions of perceptual conflict. Second, we will discuss the details of the models we developed to investigate the within-trial dynamics of the decision process in the flanker task, and the theoretical predictions of each. Third, we present the methods and results of the three experiments that served as a testbed for our model investigation, as well as the details of our model-fitting procedures. Lastly, we provide an interpretation of our results and a discussion of our findings.

3.2 Model development

Two existing SSMs of the flanker task were central to our investigation: the shrinking spotlight model of White and colleagues (2011) and the dual-stage, two-process model of Hübner and colleagues (2010). Given our specific interest in within-trial mechanisms of attention, we selected these models due to their intended fidelity to findings from the attention literature. Both models were designed as variants of the *diffusion decision model* (DDM), in which a single accumulator accrues evidence through time toward one of two response boundaries (Laming, 1968; Ratcliff, 1978). The single-accumulator structure is meant to represent the difference in firing between populations of neurons tuned to each choice (P. Smith & Ratcliff, 2004). While the standard DDM assumes evidence accumulation proceeds at a constant drift rate through time, the SSP and DSTP include alternate implementations of a time-varying drift rate in order to capture conditional accuracy effects in the flanker task.

The SSP follows the *zoom lens metaphor of attention*, in which attention is represented by a gradient of strength about a central focal point that can expand and contract alongside the area of the visual field. Retinotopic mapping studies in fMRI have provided evidence that visual attention is indeed oriented around a central fixation point in a graded fashion (Brefczynski & DeYoe, 1999; Tootell et al., 1998) and that attention-related neural activity negatively scales with the size of the attended region in a zoom lens-like manner (N. Müller et al., 2003). This work contributed to the idea that attentional resources are finite, and that top-down selective processing is necessary for preferentially allocating attention to behaviorally-relevant stimuli and events (Mesulam, 1990, 1999). In the SSP, the spotlight concept is implemented as a density function for a Gaussian distribution that is centered on the target, and each item (e.g. arrow) in

the stimulus occupies one unit of perceptual space. The standard deviation of the spotlight shrinks as a function of time, and drift rate is calculated at each time step based on the area under the curve allocated to each item. Though attempts to fit the SSP to data from tasks other than the flanker task have yielded mixed results (Servant et al., 2014; Ulrich et al., 2015), the model is still able to capture a wide range of behaviors across task conditions (White et al., 2011) and includes recoverable parameters governing the time-varying drift rate (White et al., 2018).

The DSTP, in contrast, builds off of the *dual-process hypothesis*, which proposes that two processing routes take effect when a stimulus appears: a direct, automatic route dominated by the perceptual qualities of the stimuli, and a slower, effortfully-controlled route that depends on the goal at hand (De Jong et al., 1994; Kornblum et al., 1990). As illustrated by Figure 3.2, the DSTP specifies two discrete stages of visual processing: 1) an early stage for identifying simple stimulus features and perceptual filtering, and 2) a late stage dedicated to processing the target. The early stage is divided into two racing diffusion processes: a stimulus selection phase and a response selection phase. Boundaries in each phase represent target and flanker stimuli, respectively. If the response selection phase terminates first, a response corresponding to the crossed boundary is made immediately, based only on the perceptual features of the stimulus. If the stimulus selection phase terminates first, the model transitions into the late, target-processing stage (stage 2). In Stage 2, the drift rate of the response selection phase shifts to reflect the outcome of the stimulus selection phase. The starting value of Stage 2 equals the value of the response selection process at the time that the stimulus selection process crossed a boundary. The direction of the drift rate in Stage 2 reflects the choice outcome of the stimulus selection phase. While this model can capture behavioral data patterns on a flanker task under various conditions

and has gained support from electromyography data (Servant et al., 2015), a recent parameter recovery study indicated that the drift rate parameters could not be reliably recovered from simulated data (White et al., 2018).

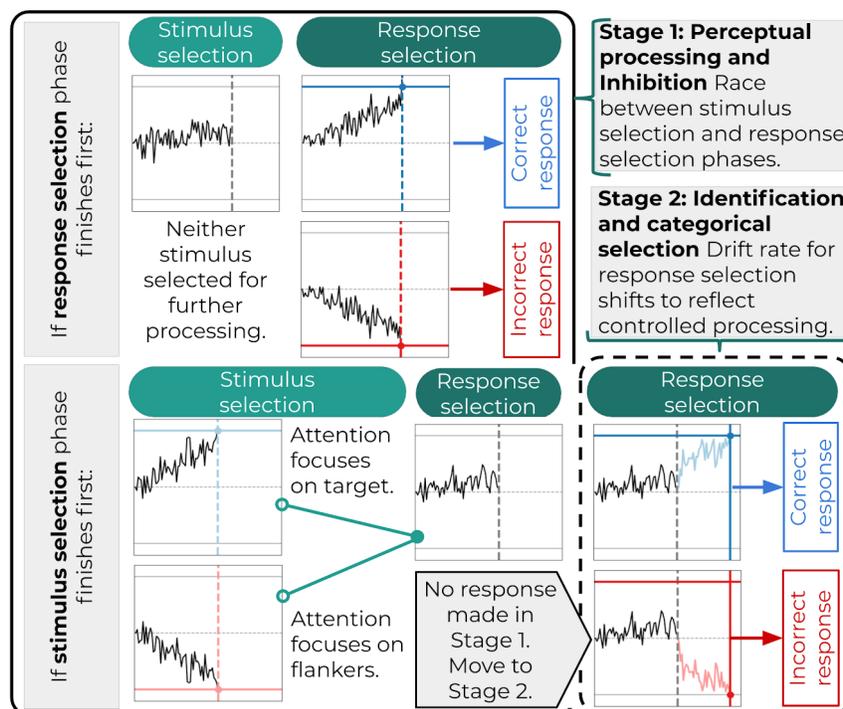


Figure 3.2: Diagram of the dual-stage two-phase (DSTP) model. In Stage 1 (left panel), the stimulus selection and response selection phases are represented by racing diffusion processes. If the response selection phase finishes first, a response is made based only on the dominant perceptual features in the stimulus array. If the stimulus selection phase finishes first, no response is made in Stage 1, and either the target or the flankers are selected for controlled attentional processing. In Stage 2 (right panel), the response selection phase drift rate changes to reflect the outcome of the stimulus selection phase.

We selected the SSP as the basis of our model investigation, systematically modifying the original model to incorporate an attentional spotlight driven by cognitive control as well as strongly- and weakly-correlated evidence accumulation mechanisms. The continuous, single-process format of the SSP was amenable to these modifications, whereas the multi-step architecture of the DSTP imposes constraints on when perceptual conflict can occur during a decision. Within our comparison of model mechanisms, our goal was to test the theory that cognitive control and related modulation of attention are emergent properties of the dynamics of the decision process. Our hypothesis, as implemented in the SSP framework, assumes that these dynamic processes interact and update continuously throughout a trial. While the cognitive control processes in the DSTP are generally time-based because the stimulus selection phase is a diffusion process with a constant drift rate, one could argue that attention in the DSTP depends on decision dynamics in addition to time alone. Specifically, the switch-point in the Stage 2 response selection drift rate is determined by the outcome of Stage 1 processes, rather than occurring at a predetermined time point. We therefore fit the DSTP to the behavioral data across our three experiments in addition to our SSP variants as a point of comparison, given that the DSTP offers an alternative account of the decision-based attention processes of interest. Equations and details of our implementation of the DSTP can be found in the Appendix B. In the following sections, we provide the details of mechanisms we implemented within the SSP framework as part of our main investigation.

3.2.1 Competition between accumulators

While the original SSP was implemented within a diffusion model framework, we adapted the shrinking spotlight mechanism within a single-boundary, dual-accumulator framework. These

two classes of models make subtly different assumptions about which neural processes are represented by evidence accumulation. In the diffusion models, evidence represents the cumulative difference in firing across populations of neurons corresponding to each of two choice options. A response is made when this difference is sufficiently large, and a boundary representing one of the two choices is crossed. In contrast, evidence in the accumulator models reflects direct competition between the two most active populations of neurons during a decision. Here, a response is made when one population of neurons reaches a predetermined firing rate threshold. Models from these two classes have been fit to data and compared extensively over the past several years, with the general consensus being that different classes of models are appropriate for different kinds of decisions (P. Smith & Ratcliff, 2004). In our project, we were interested in testing which set of assumptions is appropriate for decisions involving perceptual conflict: are decisions in the flanker task based on the difference in neural representations of targets and distractors, or the active competition between them?

Evidence accumulation in our models was mathematically defined using either LCA or FFI mechanisms. LCA is a well-known example of the accumulator class of models, and was designed to reflect observed biological mechanisms in the brain (Abbot, 1991; Amit et al., 1994). Each accumulator in the LCA model passively leaks evidence through time, and is inhibited based on the strength of the other accumulators. The FFI model, in contrast, features two accumulators with crossed inputs and no leak. As in Turner et al. (2016), we constrained the FFI model so that evidence accumulation for each choice was anticorrelated with that of the other. This implementation was meant to mimic the single-accumulator diffusion model framework, in which a movement toward one decision boundary necessitated a movement away from the other.

Similarly, for the constrained FFI model, one accumulator moving toward the decision boundary requires the other to move toward zero. Figure 3.3 provides illustrations of how evidence accumulation for two choice options occurs in the FFI and LCA models. Because evidence in the constrained FFI model is anticorrelated, the path of the decision process diffuses along a single plane and the total evidence can only increase if one accumulator reaches zero, as shown in Figure 3.3, Panel C. Figure 3.3, Panel F shows that the decision path in the LCA model is not isolated to a diagonal plane due to the independence of the accumulators.

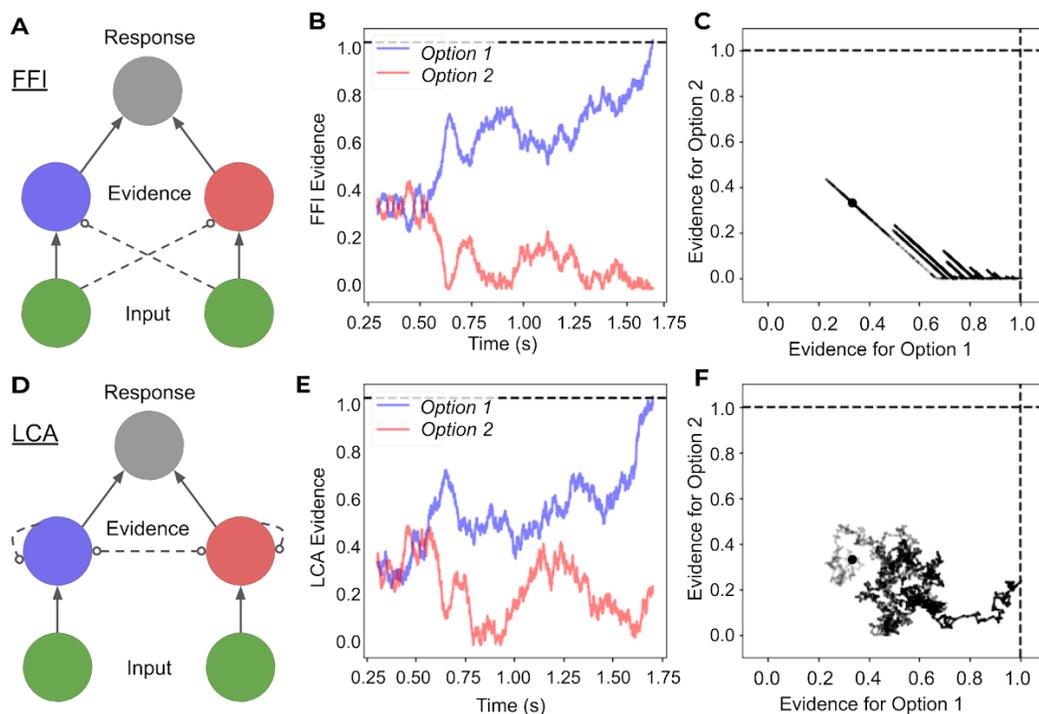


Figure 3.3: Comparison of FFI and LCA mechanisms. Left column: Graphical models of FFI (A) and LCA (D) processes from stimulus input to response, where dashed lines represent loss of evidence, open circles represent inhibition. Middle column: Simulated paths of evidence accumulation in FFI (B) and LCA (E) for two options in a single trial of a two-alternative choice

task. Right column: Phase plane plots of the same decision illustrated in panels B and E for the constrained FFI (C) and the LCA model (F). Black lines show the path of the decision process in a single trial by plotting evidence for each option against one another where 1.0 on each axis represents the decision threshold.

3.2.1.1 Constrained FFI model

Evidence for each accumulator c is denoted x_c . As described in Turner, Sederberg, & McClelland (2016), $drive_c$ and activation dx_c are represented by

$$drive_c = \rho_c \frac{dt}{\Delta t} + \xi \sqrt{\frac{dt}{\Delta t}}$$

$$dx_c = drive_c - drive_{-c}$$

$$x_c \rightarrow \max(x_c, 0).$$

where ρ_c denotes the drift rate for accumulator c . $drive_{-c}$ represents drive for the opposing accumulator with respect to c . To approximate the continuous differential equation for $drive_c$, we used the Euler method to discretize time, selecting a step size of $dt=0.01$ modified by a time constant of $\Delta t = 0.1$ (S. Brown et al., 2006). The degree of noise in the accumulation process is represented by ξ , a driftless Wiener process distributed as $\xi \sim \mathcal{N}(0, 1)$. In line with the conventions of accumulator models, evidence x_c for each accumulator c was bound at zero so that neither accumulator could ever be negative. Evidence for each alternative accumulates through time until decision threshold α is reached, and a response is selected in favor of the winning accumulator. Response time, then, is equal to the sum of the time taken for one of the accumulators to reach α and non-decision time τ , which comprises early visual processing and motor preparation. Although different approaches could have been taken, accumulator starting

points were set in relation to the decision threshold α such that $x_c = \frac{\alpha}{3}$ for $c \in \{1,2\}$. This choice of starting point has been selected in previous modeling work (Ditterich, 2010; van Ravenzwaaij et al., 2012) to align with findings from single unit recordings (Churchland et al., 2008).

3.2.1.2 LCA model

While evidence in the constrained FFI model is strongly correlated, LCA accumulators are weakly correlated, linked only by lateral inhibition processes that repel the accumulators away from one another via parameter β . Evidence for each choice passively decays throughout the accumulation process at a rate equal to leak parameter κ . Activation dx_c at is given by

$$dx_c = (\rho_c - \kappa x_c - \beta \sum_{j \neq c} x_j) \frac{dt}{\Delta t} + \xi \sqrt{\frac{dt}{\Delta t}}$$

$$x_c \rightarrow \max(x_c, 0).$$

Again, we used the Euler method to discretize time, selecting a step size of $dt=0.01$ modified by a time constant of $\Delta t = 0.1$. Evidence accumulates through time until the decision threshold α is reached, and a response is made after non-decision time τ . Evidence x_c was bound at 0 and

starting points were set to a proportion of threshold α such that $x_c = \frac{\alpha}{3}$ for $c \in \{1,2\}$.

3.2.2 Drive to attention mechanisms

Our core mechanistic hypothesis is that attention is directly modulated within-trial as an emergent property of decision-making dynamics. This hypothesis is based on evidence of within-trial changes in attention and cognitive control from neuroimaging (Czernochowski, 2015;

Nigbur et al., 2015; Scherbaum et al., 2011) and connectionist models in which cognitive control is dynamically mobilized in response to the mutual activation of multiple response nodes (De Pisapia & Braver, 2006; Frank, 2006; Scherbaum et al., 2012; Verguts, 2017). Our proposed control-driven attention mechanism stands in contrast to existing SSMs of decision processes during the flanker task, in which attention is directly dependent upon time itself. To test our hypothesis against the assumption of time-dependent attention processes, we developed variants of the SSP with either time-based or control-based attentional spotlights. The time-based models mirror the original SSP so that attention, implemented as a density function for a Gaussian distribution centered on the target of a flanker array, gradually shrinks throughout a trial as a linear function of time. In the control-based models, cognitive control is calculated as the cumulative distance between total evidence and a threshold. The standard deviation of the attentional spotlight is in turn calculated as a function of cognitive control. These mechanisms are illustrated in Figure 3.1.

Braver's DMC framework (2012) suggests that under unpredictable conditions, cognitive control (specifically reactive control) is upregulated within-trial until the available level of attention is sufficient for conflict to be successfully resolved. Cognitive control therefore serves as an interface between the state of the system and limited resources, continuously making comparisons between the *active* and *required* levels of attention and updating them accordingly. To reflect hypothesized mechanisms for continuous attentional monitoring through time, cognitive control was operationalized as the cumulative distance between total evidence and a threshold (δ). The δ threshold represents a learned level of evidence at which conflict can typically be resolved in the context of the task. At the beginning of a trial, total evidence is

maximally distant from δ , which results in the upregulation of attentional resources via cognitive control. As evidence increases throughout a trial and eventually surpasses δ , the active level of attention becomes sufficient for resolving conflict. Because no further attentional upregulation is required, cognitive control begins to decrease toward the end of the trial. Average simulations of within-trial cognitive control signals for each task condition are shown in Figure 3.4, alongside time signals for contrast. Simulations reveal higher peak levels of cognitive control on incongruent compared to congruent trials. This observation is consistent with theoretical accounts indicating that incongruent trials require more cognitive control for correct decisions than congruent (Botvinick et al., 1999; Gratton et al., 1992). In summary, we developed a measure of cognitive control that is based on the dynamics of the evidence accumulation process, generally builds through time, is able to relax toward the end of a trial, and naturally demonstrates differences between task conditions.

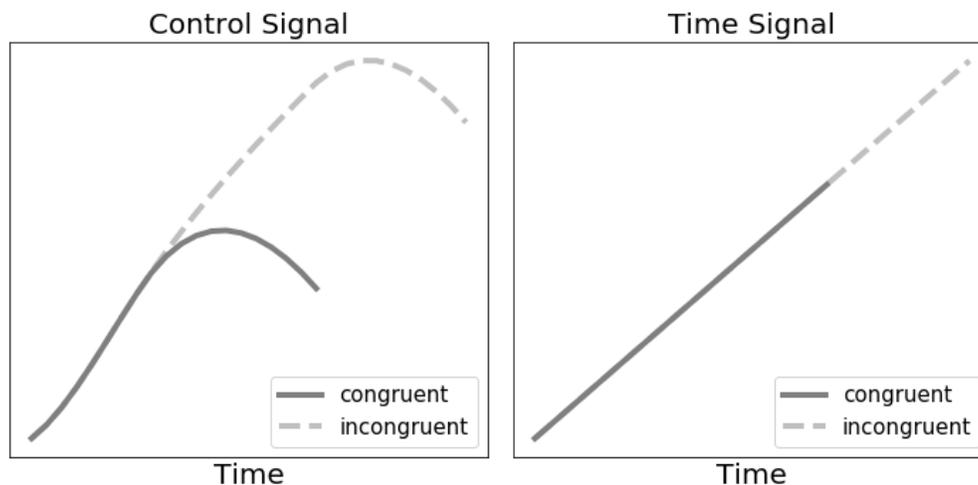


Figure 3.4: Control- and time-based signals to attention. Across models in our comparison, attentional spotlights shrink as a function of control (Left panel) or time (Right panel). Mean simulations of control and time signals are shown for a single trial in the congruent and incongruent conditions.

Because our calculation of cognitive control is based on the sum of evidence at each time step, the mode of evidence accumulation (FFI vs. LCA) has notable effects on the moment-to-moment changes in cognitive control, and subsequently, the behavior of the spotlight. The accumulators in the FFI model are strongly correlated as shown by the phase plane plots in Figure 3.3, and total evidence only changes if one accumulator is forced to zero while the other continues to increase. Otherwise, an increase in evidence for one accumulator results in a decrease in evidence for the other, and the sum of evidence remains constant. For weakly-correlated LCA accumulators, however, total evidence fluctuates as rapidly as the accumulator values themselves. While spotlights in both FFI-control and LCA-control models share the general characteristics of narrowing through time at variable rates while maintaining the ability to widen as cognitive control relaxes, LCA-control naturally predicts a spotlight trajectory with higher

within-trial variability in comparison to FFI-control. Due to the possibility that noise alone would result in similarly-fitting models compared to the mechanisms of interest, we developed FFI and LCA model variants in which the spotlight is driven by time with additional within-trial variability. The standard deviation of the noise distribution was added as a free parameter, so that variability in the spotlight calculation could be added as needed to optimally fit the data. Figure 3.5 shows calculations of attentional spotlight widths through time, generated from the FFI-conflict and LCA-conflict models as well as time- and time+noise-based models. In the following sections, we will provide the mathematical details of each type of attentional spotlight mechanism that we explored in the current project.

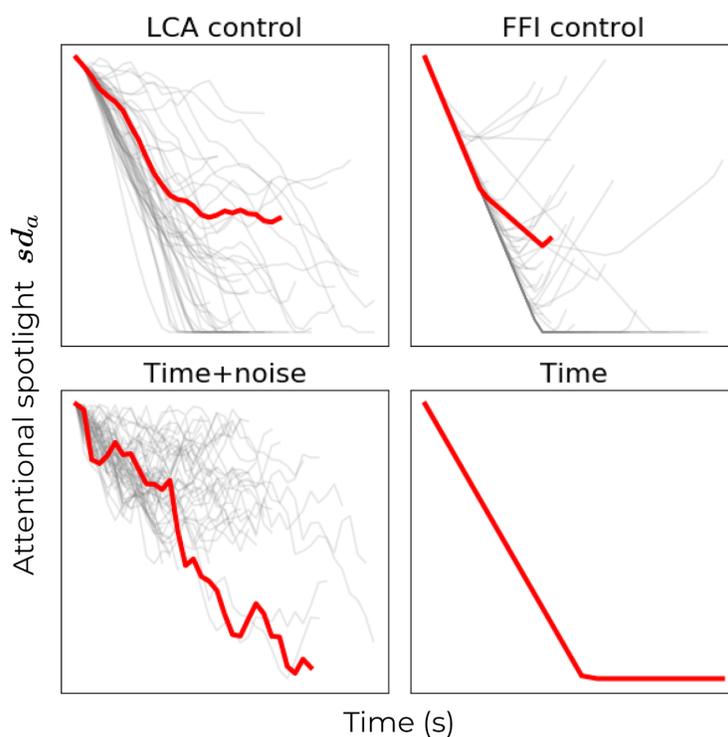


Figure 3.5: Model-generated spotlight widths through time. For each model, 50 trials were simulated from one participant's best-fitting parameters. Panels show calculations of spotlight standard deviations through time, with each simulation displayed as a gray line to demonstrate

the between-trial variability captured by each model. A single additional simulation is shown as a red line to illustrate differences in within-trial variability.

3.2.2.1 Time-based attention

As in the original SSP, our two-accumulator implementations of the model calculate drift rate through time based on an attentional spotlight. Drift rate is governed by three free parameters: attentional strength (p), width of the spotlight at the beginning of a trial (sd_0), and the rate at which the spotlight shrinks (r_d). Across models, the spotlight is a density function for a Gaussian distribution centered at 0 with standard deviation (sd_a). The width of the spotlight is calculated continuously as a function of time, discretized as t :

$$sd_a = sd_0 - r_d t$$

Equation 3.1

and the area of the attended spatial region allocated to target and flanker items is given by

$$a_{target} = \int_{-0.5}^{0.5} \mathcal{N}(0, sd_a)$$

$$a_{flanker} = \int_{0.5}^{n+0.5} \mathcal{N}(0, sd_a)$$

where n is the number of flanker items on each side of the target on a horizontal plane.

Allocation of spatial attention based on the area under a Gaussian curve is illustrated in Panel 2 of Figure 3.1. Limits reflect the assumption that each item in the stimulus array occupies one unit of perceptual space (White et al., 2011). Drift rates for the correct (ρ_2) and incorrect (ρ_1)

responses are calculated in each condition depending on the direction of the flanker items relative to the target via

$$\text{congruent} : \rho_2 = pa_{target} + 2pa_{flanker}; \rho_1 = 0$$

Equation 3.2

$$\text{incongruent} : \rho_2 = pa_{target}; \rho_1 = 2pa_{flanker}.$$

Equation 3.3

3.2.2.2 Control-based attention

In contrast to time being the driving force to the attentional spotlight, we defined a subset of models in which the spotlight standard deviation was calculated continuously as

$$sd_a = sd_0 - r_d c$$

Equation 3.4

where c represents cognitive control. As described previously, cognitive control was calculated based on the cumulative distance between the total amount of evidence in the system and a conflict resolution threshold δ , such that

$$dc = \left(\delta - \sum_{c \in \{1,2\}} x_c \right) \frac{dt}{\Delta t}$$

As in the time models, drift rates were calculated via Equations 3.2 and 3.3.

3.2.2.3 Time with noise

As shown in Figure 3.5, the control-based models allow for more variability in drive to the attention system compared to the time models. While this variability is a natural consequence of

calculating sd_a based on the state of noisy accumulators, we wanted to investigate whether the addition of *random* variability would be equally suitable for fitting the data. As such, we developed variants of the time models that included an additional free parameter σ . Noise ζ was drawn from a driftless Wiener process such that $\zeta \sim \mathcal{N}(0, 1)$. The standard deviation of the spotlight was then calculated from the noisy time-based signal, such that

$$d\eta = \sigma \zeta \frac{dt}{\Delta t}$$

$$sd_a = sd_0 - r_d \eta.$$

3.2.3 Summary of model variants

Our current investigation was centered around four variants of the SSP, each containing a different combination of evidence accumulation mechanisms (strongly-correlated, FFI vs. weakly-correlated, LCA) and calculations for visual attention (time-based vs. control-based). Because the control-based models allow for variability in the behavior of the attentional spotlight whereas the time-based models do not, we included FFI and LCA variants of time models in which within-trial noise was injected into the spotlight calculation. Table 3.1 summarizes the free parameters included in each of these six models. To investigate an alternative method for decision-based mechanisms for attention and cognitive control, we also included the DSTP model. The 9 free parameters in the DSTP model are listed in Appendix B.

Table 3.1: Summary of free parameters across models in Chapter 3

| Parameter | Description | Model | | | | | |
|--------------|---------------------------|-------------|-------------------|----------------|-------------|-------------------|----------------|
| | | FFI time | FFI time+noise | FFI control | LCA time | LCA time+noise | LCA control |
| r_d | rate of focus | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| p | perceptual input strength | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| sd_0 | starting spotlight width | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| α | decision threshold | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| τ | non decision time | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| σ | within-trial variability | | ✓ | | | ✓ | |
| δ | conflict threshold | | | ✓ | | | ✓ |
| κ | leak | | | | ✓ | ✓ | ✓ |
| β | lateral inhibition | | | | ✓ | ✓ | ✓ |
| Total | | 5 | 6 | 6 | 7 | 8 | 8 |

3.3 Experiments

Data from three experiments served as the testbed for the seven model variants. The first experiment was a standard flanker task, which was intended to test each model's ability to capture basic behavioral effects between conditions. The second experiment included a manipulation in which the perceptual strength of the target relative to the flanker items varied from trial to trial. These data were fit by adding free parameters to modify perceptual input strength (p) depending on the perceptual strength of each item in the stimulus array. The third experiment was a standard flanker task during which we recorded scalp EEG measurements. The models were fit to behavior alone for all experiments, and simulation methods were used in our analysis of data collected in Experiment 3 to observe which models most successfully mapped onto within-trial EEG voltage at each electrode.

3.3.1 Experiment 1

Given that the SSP was designed to capture data in a standard flanker task and has successfully fit patterns of responses across conditions (White et al., 2011), we wanted to test all of our SSP model variants in this domain as well. Participants completed a standard flanker experiment, in which they indicated the direction of a central arrow while ignoring congruent, incongruent, or neutral distractor items. Although we only fit the models to data from congruent and incongruent trials, we hoped that the inclusion of neutral trials would boost flanker effects via increased rarity of incongruent trials (Gratton et al., 1992) while maintaining equal numbers of congruent and incongruent observations.

3.3.1.1 Methods

3.3.1.1.1 Procedure

After providing written informed consent, participants were seated in a cubicle and asked to turn off all electronic devices. Instructions for the task appeared on the computer screen, and were read aloud by the experimenter. Each block began with a summarized instruction screen to remind participants of the appropriate response mappings while also providing an opportunity to take a short break from the task. The instruction summary remained on the screen until the participant pressed the ENTER key to proceed. During each trial, a fixation cross appeared in the center of the screen for 1000 ms before being removed. The trial stimulus then appeared on the screen after a jittered duration of 100-900 ms. Participants responded by pressing the 'J' key on the keyboard if the arrow in the center of the array pointed left, and the 'K' key if the center arrow pointed right. Participants were asked to respond with their right forefinger and right middle finger respectively. Only responses made 150 ms after the stimulus appeared were

recorded, and the stimulus was removed from the screen immediately after the participant made a valid response. Participants were given an unlimited amount of time to respond, but were instructed to respond as quickly and accurately as possible.

3.3.1.1.2 Stimuli and apparatus

A custom program using the State Machine Interface Library for Experiments (SMILE; <https://github.com/compmem/smile>) was written to present stimuli, track timing, and log responses. Stimuli were presented on a desktop computer equipped with Linux OS connected to a 15-inch display with a refresh rate of 60 Hz. Participants were seated in individual cubicles within view of an experimenter. Before beginning, participants completed 10 practice trials of the task. The task consisted of 8 blocks of a standard flanker task, each block containing 48 trials. Including practice, participants completed 394 trials in total. Task condition (congruent, incongruent, neutral) and target direction (left, right) were counterbalanced within block. Stimuli were presented in white font on the horizontal midline of a dark gray field. Each stimulus consisted of a target arrow in the center of 6 flanker items, 3 to the left and 3 to the right.

3.3.1.1.3 Participants

27 undergraduate students at The Ohio State University participated in Experiment 1 in exchange for partial course credit. All participants provided informed consent in accordance with the requirements of the Institutional Review Board at the university. One participant's data were excluded from analysis due to failure to exceed a chance level of performance on the task.

3.3.1.1.4 Model-fitting and comparison

The seven models were fit to each participant's data independently using probability density approximation (PDA) methods described by Turner and Sederberg (2014) and implemented via custom programs with RunDEMC (<https://github.com/compmem/RunDEMC>). Because the models within the current investigation do not have analytic likelihood functions, PDA methods allowed us to approximate how likely the choice and RT data Y would be under a set of model parameters θ . After specifying each model, we defined a set of prior distributions $\pi(\theta)$ for each parameter that will be discussed in the next section. Parameter sets were proposed via differential evolution with Markov chain Monte Carlo (DE-MCMC; Ter Braak, 2006; Turner et al., 2013; Turner & Sederberg, 2012), a genetic algorithm that makes proposals based on the relative success of previous proposals. Within DE-MCMC, a proposed parameter set in a chain is accepted with Metropolis Hastings probability, such that parameters have a higher probability of survival if they fit the data better than the previous proposal, and concurrent chains inform one another on each iteration. Using each proposed parameter set θ^* , we simulated each model 30,000 times to produce a set of data X such that $X \sim \text{Model}(\theta^*)$. From these distributions, we constructed a simulated probability density function using an Epanechnikov kernel (Turner et al., 2016; Turner & Sederberg, 2014) to estimate the form of X . We then calculated the density of each point in the observed data Y under the given set of parameters θ using the equation:

$$\text{Model}(Y_i|\theta) = f(Y_i|X)$$

Where f is an approximation of the functional form of simulated data X . We then approximated the likelihood function using the equation

$$L(\theta|Y) = \prod_{i=1}^N \text{Model}(Y_i|\theta)$$

Finally, the posterior density for a given parameter set was approximated by combining the likelihood function and the set of prior distributions $\pi(\theta)$ with the equation:

$$\pi(\theta|Y) \propto L(\theta|Y)\pi(\theta).$$

This procedure was implemented in 50 chains for 600 “burn-in” iterations to identify the maximum a posteriori (MAP) estimate, followed by 1,600 sampling iterations to generate full posterior distributions. A purification step was implemented every 5 iterations for the accepted population, in which likelihood values were recalculated and replaced in order to prevent chains from getting stuck in spuriously high-likelihood regions of the posterior (Holmes, 2015; Turner et al., 2018). Priors were selected to be uninformative in terms of range, but to provide a moderate level of constraint in terms of functional form. As none of these models have been fit in a Bayesian paradigm, we had no precedent to rely upon for selecting a prior distribution for each parameter. Prior distributions were specified as follows, and were the same across models that utilized common parameters:

$$r_d \sim \mathcal{U}(0, 20)$$

$$p \sim \mathcal{U}(0, 20)$$

$$sd_0 \sim \mathcal{TN}(1, 10, 0, 20)$$

$$\alpha \sim \mathcal{TN}(2.5, 10, 0, 30)$$

$$\tau \sim \mathcal{TN}(0.1, 0.5, 0, \min(RT))$$

$$\text{logit}(\kappa) \sim \mathcal{N}(0, 1.4)$$

$$\text{logit}(\beta) \sim \mathcal{N}(0, 1.4)$$

$$\delta \sim \mathcal{TN}(2.5, 10, 0, 30)$$

$$\sigma \sim C^+(0, 10)$$

To compare the relative fit performances of the models, we calculated the Bayesian predictive information criterion (BPIC; Ando, 2007) for each model within-subject. We selected BPIC as our comparison metric for the present investigation because it is calculated in consideration of the full posterior distribution rather than a point estimate of the maximum log likelihood. This metric also accounts for model complexity by favoring models with fewer free parameters. To calculate BPIC values, a vector $V(\theta)$ of deviance values was calculated from the likelihood θ for each set of parameters in the latter 1,400 sampling iterations of the posterior using the equation:

$$V(\theta) = -2\log(L(\theta|D)).$$

We then calculated the mean and minimum deviance as \bar{V} and \hat{V} respectively. The effective number of parameters p_V was calculated as $p_V = \bar{V} - \hat{V}$. Finally, the BPIC value was calculated as:

$$BPIC = \bar{V} + 2p_V$$

where lower BPIC values indicated a better fit.

3.3.1.2 Results

3.3.1.2.1 Behavior

Responses shorter than 150 ms or longer than 2000 ms were excluded from analyses and model-fitting (<4% of trials across subjects). Neutral trials were excluded from analyses due to an unforeseen pop-out effect in our data, such that participants were slightly faster at responding to neutral stimuli compared to congruent. Only congruent and incongruent trials were analyzed further. A summary of behavioral results is shown in Table 3.2. Behavioral results were analyzed using paired-sample t-tests, where the degrees of freedom for within-condition performance comparisons were based on the number of subjects who made at least one error in the condition

of interest. We observed the expected flanker task effects, including significantly lower accuracy on incongruent trials compared to congruent ($t(25)=-2.919$, $p<0.01$) and significantly slower RTs for incongruent trials compared to congruent ($t(25)=7.520$, $p<0.001$). Our data also demonstrated significantly faster errors than correct responses in the incongruent condition ($t(22)=-3.778$, $p<0.01$), but not in the congruent condition ($t(9)=0.910$, $p=0.386$).

Table 3.2: Average accuracy and mean RTs (ms) across participants for Chapter 3, Experiment 1

| Condition | Accuracy | All RT | Correct RT | Error RT |
|-------------|----------|--------|------------|----------|
| Incongruent | 0.912 | 661 | 669 | 533 |
| Congruent | 0.969 | 537 | 540 | 620 |

3.3.1.2.2 Model fits

BPIC values were calculated for each model and subject as a measure of goodness-of-fit. Values were mean-centered within subject, and are displayed as a heat map in Figure 3.6. Out of 26 total participants in Experiment 1, the LCA control model was the best performing model for 8 participants, the FFI control model was the best performing model for 5 participants, the LCA time model was the best performing model for 2 participants, the DSTP model was the best performing model for 7 participants, and the FFI time model was the best performing model for 4 participants. Accounting for the magnitude of the wins across subjects, the two conflict models outperformed their time-based alternatives and DSTP, though results were mixed overall.

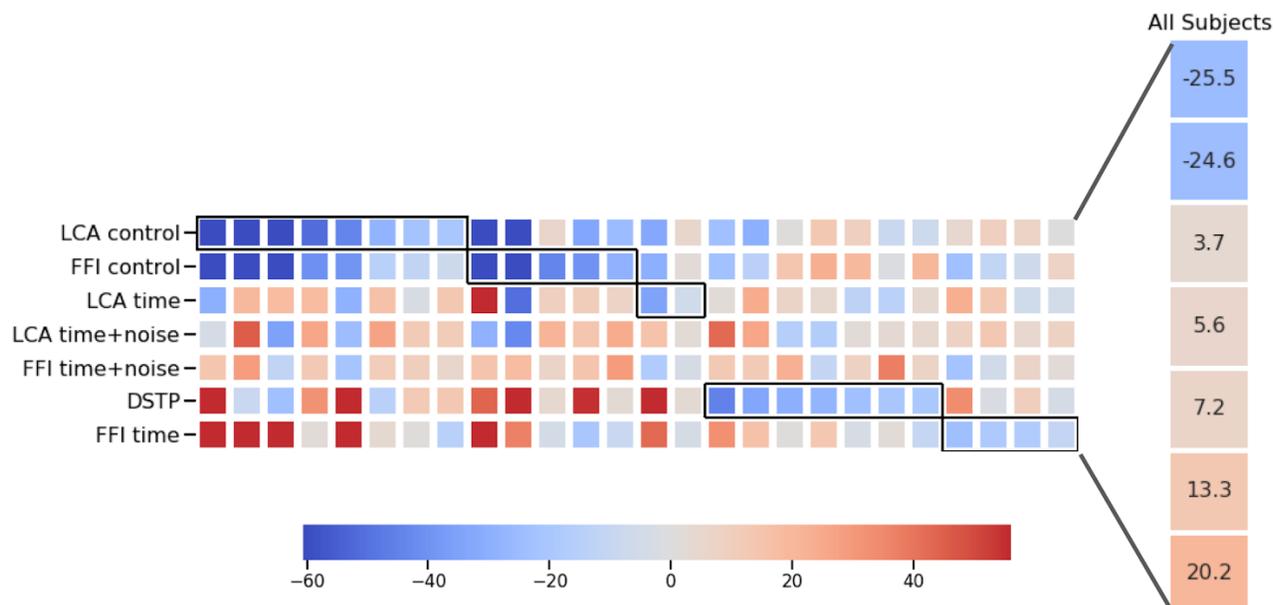


Figure 3.6: Heat map of BPIC values, mean-centered within-subject for Experiment 1. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right.

Figure 3.7 shows observed choice-RT distributions averaged across participants, as well as mean distributions generated from each subject's best-fitting parameters in our four main models of interest. All four models were able to capture typical flanker effects of slower, less accurate responses in the incongruent compared to the congruent condition, and faster errors than correct responses in the incongruent condition. Given that the SSP was specifically designed to capture robust congruency and conditional accuracy effects, it is unsurprising that all models were able to fit the standard pattern of data. Though the control-driven models were better suited for capturing the peaks of the correct response distributions than the time-driven models, across-subject results reflect strong model mimicry. To gain more insight into the differences in

predictions among the models, we need to delve into the more nuanced patterns of behavior that were not necessarily robust across all subjects.

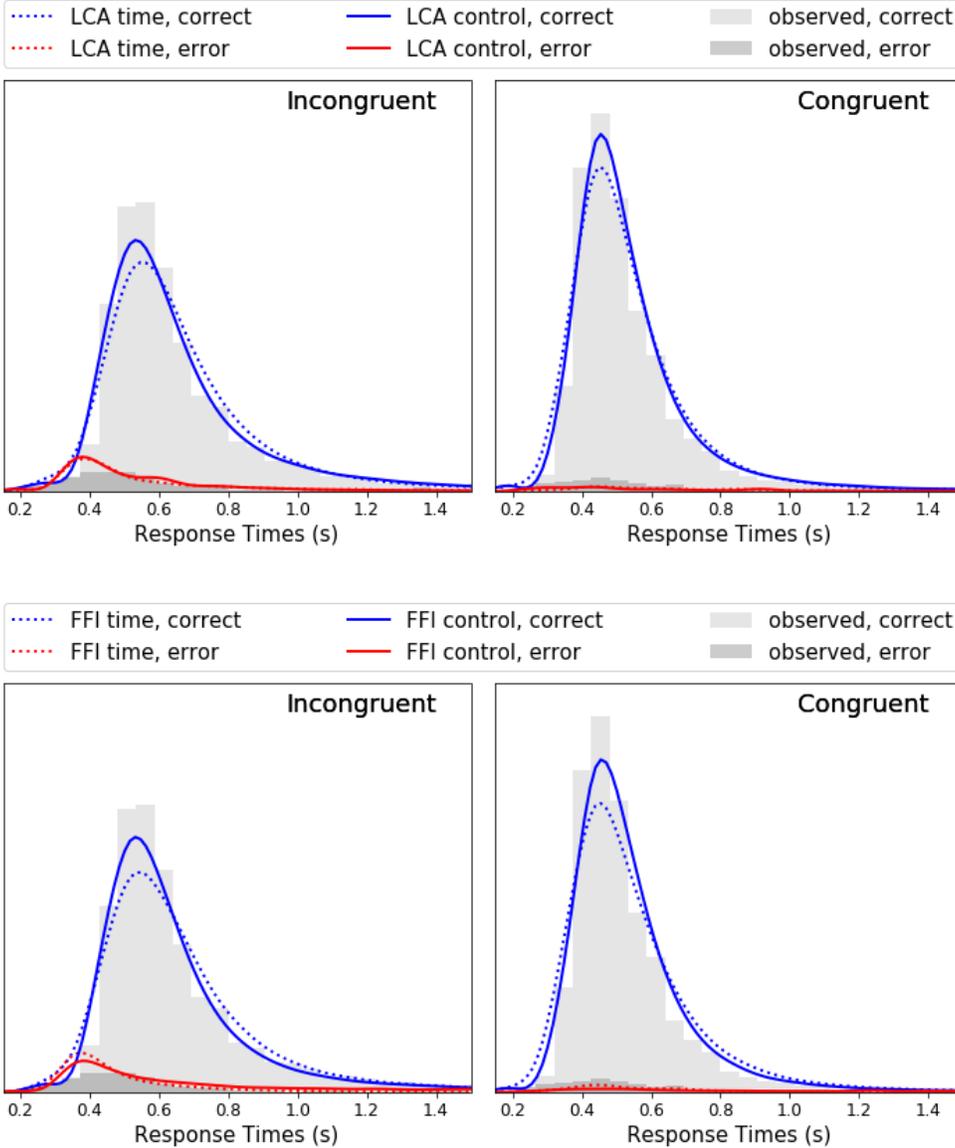


Figure 3.7: Observed and model-generated choice-RT distributions. Observed RT distributions for correct (light gray histograms) and incorrect (dark gray histograms) responses were averaged across participants. Models were simulated 10,000 times for each condition, using each

participant's best-fitting parameters. Lines show average model-generated distributions across participants. Distributions generated by the FFI time and FFI control models are shown in the top row, whereas distributions generated by the LCA time and LCA control models are shown in the bottom row.

We provide analyses using two measures of response capture: *error location indices* (ELIs) and *conditional accuracy functions* (CAFs). An ELI value represents the proportion of incorrect responses that are faster than trials chosen at random (Servant et al., 2018). For example, a participant who performed less accurately when they made fast responses would likely have a high (close to 1.0) ELI, whereas a participant who performed less accurately when they made slower responses would likely have a low (close to 0.0) ELI. The SSP was developed to capture the general effect of fast errors specific to the incongruent condition of the flanker task, which manifests as higher ELI values in the incongruent compared to the congruent condition. While all four of the main models in the current investigation can capture this basic effect, we observed differences among the models in terms of their abilities to predict individual differences in ELIs in the incongruent condition. After fitting each model to data from each subject, we used best-fitting parameters to generate predicted ELI values. Figure 3.8 shows correlations between observed and predicted ELI values in the incongruent condition for each model. Per the requirements of the calculation, participants were only included if they made at least one error in the incongruent task condition (23 participants). These results suggest that the LCA control model is best able to capture the nuanced subject-level differences that we observed in our dataset. To assess significance, we applied a Fisher's z transformation to each r correlation and calculated an observed z test statistic at an alpha level of 0.05 for each pairwise combination of

models. The observed vs. predicted ELI correlation for the FFI control model was significantly lower than that of the LCA control ($z=2.284$, $p=0.011$) and LCA time models ($z=1.742$, $p=0.041$). No other comparisons were significant (LCA control vs. LCA time: $z=0.542$, $p=0.294$; LCA control vs. FFI time: $z=0.833$, $p=0.203$; LCA time vs. FFI time: $z=0.291$, $p=0.386$; FFI time vs. FFI control: $z=1.451$, $p=0.073$).

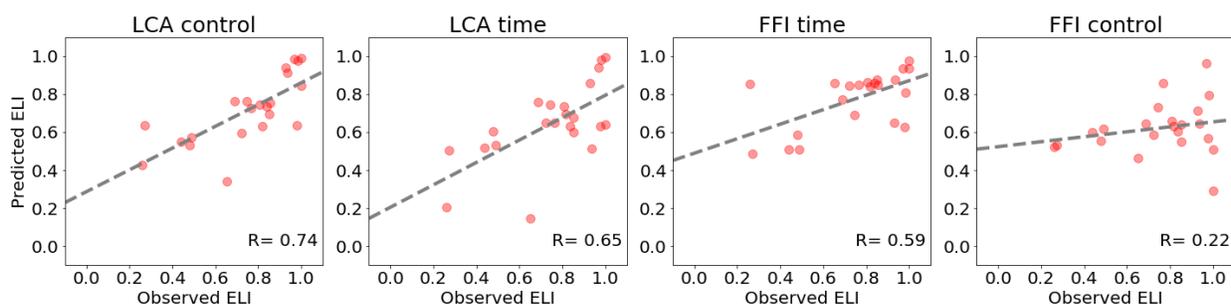


Figure 3.8: Observed and predicted ELI values for incongruent trials. ELI values calculated from each subject's data in the incongruent condition (x-axis) are plotted against the ELI values generated from each subject's best-fitting parameters (y-axis) in each model (panels).

Correlations and lines of best fit are displayed on each panel.

While ELIs were developed as a quantitatively interpretable alternative to CAFs, CAFs remain a common tool for illustrating behavioral effects in the flanker task. In the CAF, performance is plotted as a function of RT. Figure 3.9 shows average CAFs across subjects calculated from observed data in the incongruent condition, overlaid by average predicted incongruent CAFs generated from each subject's best-fitting parameters for each model. As mentioned previously, all four models can capture fast errors in the incongruent condition, which is illustrated by lower accuracy in the initial RT bins. The models differ, however, in their abilities to capture *slow*

errors. Neither the LCA time nor the FFI time model appropriately captures the observed drop-off in accuracy for longer RTs. The control models, however, are able to predict a decrease in cognitive control toward the end of a trial, which allows the models to capture patterns of accuracy that reach a peak before slightly decreasing. This is due to the nature of our conflict signal as illustrated by Figures 3.1 and 3.4, which allows for the widening of the attentional spotlight toward the end of a trial depending on the parameter values. The FFI control model, however, appears to overpredict the proportion of slow errors due the combination of the control mechanism and the strong correlation between accumulators, resulting in the lowest correlation between observed and predicted ELI values across the models as shown in Figure 3.8.

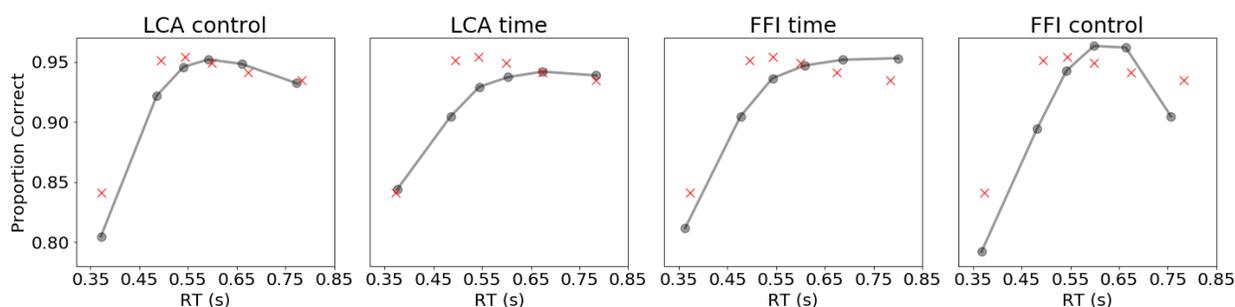


Figure 3.9: Observed and predicted CAFs for incongruent trials. Data from each subject were sorted according to RT within 6 equally-spaced percentile bins. Performance and minimum RT for each bin were averaged across participants (red Xs). After generating 1,000 choice-RT pairs from each subject's best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines).

ELIs for the congruent condition were useful for distinguishing these models as well. Similar to Figure 3.8, Figure 3.10 shows ELI values calculated from observed data in the congruent

condition in relation to the predicted ELI values generated from best-fitting parameters in each model. Per the requirements of the calculation, participants were only included if they made at least one error in the congruent task condition (10 participants). Predictions using the LCA control model best mapped onto subject-level ELIs in the congruent condition compared to the other models. The observed vs. predicted ELI correlation for the LCA control model was significantly higher than that of the FFI time ($z=3.088$, $p=0.001$) and FFI control models ($z=1.871$, $p=0.031$), and the correlation for the LCA time model was significantly higher than that of the FFI time model as well ($z=1.822$, $p=0.034$). No other comparisons were significant (LCA control vs. LCA time: $z=1.266$, $p=0.103$; LCA time vs. FFI control: $z=0.606$, $p=0.272$; FFI control vs. FFI time: $z=1.217$, $p=0.112$).

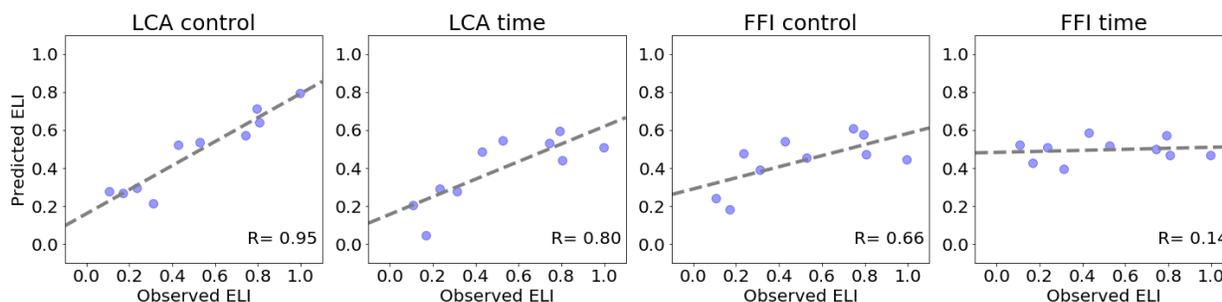


Figure 3.10: Observed and predicted ELI values for congruent trials. ELI values calculated from each subject's data in the congruent condition (x-axis) are plotted against the ELI values generated from each subject's best-fitting parameters (y-axis) in each model (panels). Correlations and lines of best fit are displayed on each panel.

To observe specific differences in model predictions within the congruent condition, mean CAFs were generated separately for participants with low (0.11-0.31) and high (0.74-1.00) observed

ELIs as determined by median split. Figure 3.11 shows observed and model-predicted CAFs for low-ELI participants in the congruent condition, in which the observed data demonstrates a higher proportion of errors for longer compared to shorter RTs. While all models miss the mean performance values considerably, the LCA control, LCA time, and FFI control models are able to capture a general pattern of slow errors in the congruent condition. Though the LCA time model lacks the ability to relax attentional processing like the control models, it is presumably able to capture these slow errors via the leak (κ) parameter. The FFI time model, however, has no mechanism for capturing slow errors in the congruent condition.

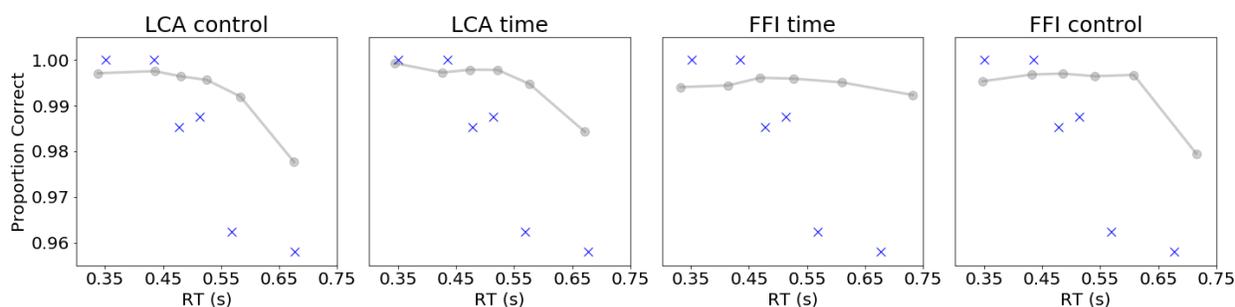


Figure 3.11: Observed and predicted CAFs for congruent trials across low-ELI participants. Data from each subject were sorted according to RT within 6 equally-spaced percentile bins.

Performance and minimum RT for each bin were averaged across participants (blue Xs). After generating 1,000 choice-RT pairs from each subject's best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines).

Figure 3.12 shows observed and predicted CAFs for high-ELI participants in the congruent condition. The observed data demonstrates a higher proportion of errors for shorter compared to longer RTs. Neither of the time models are able to predict fast errors in the congruent condition.

While the cognitive control-driven attentional system allows the FFI control model to predict fast errors, these processes in combination with a strongly-correlated accumulator structure result in an overprediction of slow errors. The LCA control model, however, is able to predict fast errors without inappropriately predicting slow errors as well.

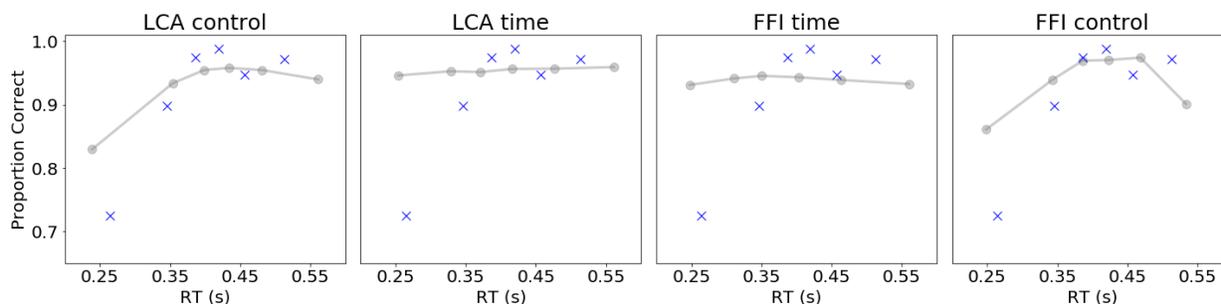


Figure 3.12: Observed and predicted CAFs for congruent trials across high-ELI participants.

Data from each subject were sorted according to RT within 6 equally-spaced percentile bins.

Performance and minimum RT for each bin were averaged across participants (blue Xs). After

generating 1,000 choice-RT pairs from each subject's best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines).

3.3.1.3 Discussion

The results of Experiment 1 demonstrate strong mimicry between models, but showed overall

better fits for models with control-driven attentional mechanisms compared to time-driven

alternatives as determined by our BPIC comparison. In interpreting the BPIC results, it is

important to remember that these calculations favor less complex models. With 8 free

parameters, it is therefore notable that the LCA control model outperformed the more

parsimonious alternatives in a substantial number of cases. For the 4 instances in which the more

parsimonious FFI time model was the best-fitting model, it appears that the improvements in fit afforded by the more flexible models were not substantial enough to justify the additional complexity. The *most* complex model was the DSTP with 9 free parameters, and its flexibility resulted in best fits for 7 subjects. For a majority of subjects, however, the added complexity did not improve the fits over what the other models could provide, and the model barely performed better than FFI time on average. Interestingly, the control models provided better fits than the time+noise models in almost all cases, indicating that the control mechanisms themselves are tapping into an aspect of the data beyond improvements resulting from additional noise. Because each model makes the standard predictions for choice-RT distributions equally well, ELI and CAF analyses allowed us to investigate the predictions of the models at a finer granularity than what choice-RT summarizations could provide. Among the FFI time, LCA time, FFI control, and LCA control models, only LCA control could predict patterns of fast and slow errors in each condition that varied by subject. Although Experiment 1 has provided tentative evidence that cognitive control, rather than time alone, underlies attention processes in the flanker task, the data as a whole did not provide a strong dissociation between FFI and LCA mechanisms for interactions between the accumulators when considering general effects across subjects. To investigate the possibility that either β or κ was independently responsible for the success of the LCA control model above FFI control, we fit variants of LCA control to data from Experiment 1 in which β and κ were fixed to 0, respectively. Although both models outperformed the FFI time model as determined by BPIC, neither of these variants provided better fits than the LCA control with free β and κ . Detailed results are provided in Appendix B. We also wanted to see if the success of the LCA control model over the LCA time model was due to the use of a nonlinear signal for guiding the attentional spotlight. Although a linear definition of time was adopted

within the original SSP model, our results with the control-based models suggested that a sublinear function for governing the spotlight might provide better fits to data. We developed a variant of the LCA time model (“LCA exponential time”) in which time was an exponential function increasing up to an asymptote, which was added as a free parameter. In a BPIC comparison shown Appendix B, this model fit worse than LCA control for all 26 subjects in Experiment 1.

3.3.2 Experiment 2

Because the results of Experiment 1 did not favor strongly-correlated FFI or weakly-correlated LCA evidence accumulation mechanisms, we next fit the models to data from a task that we believed would challenge these alternative hypotheses. In the standard flanker task, the nature of the arrow stimuli results in an equal amount of perceptual strength for each item in an array, and evidence for a left response is equal and opposite to evidence for a right response. As such, it is not surprising that both FFI and LCA accumulation dynamics were able to capture the data equally well. In Experiment 2, we opted to test the models under task conditions in which the perceptual strength of the flanker items was not necessarily equal to that of the target. The task, designed and administered by Servant et al. (2014), required participants to indicate the color of a target circle amid flanker circles of a congruent or incongruent color. As a manipulation of relative perceptual strength, the color saturation of the target circle varied from trial to trial while the saturation of flanker circles was held constant. Due to the strongly-correlated behavior of the accumulators in the FFI models, we predicted that the FFI models would be less capable of capturing the observed patterns of choices and RTs across conditions in this task relative to LCA models. Our hypothesis is in line with recent work showing that models with strongly-correlated

accumulators fail to capture observed patterns of data across a range of equal- and unequal-evidence task conditions (Kirkpatrick et al., *under review*).

3.3.2.1 Methods

3.3.2.1.1 Procedure

The data set used in the present investigation was collected at Aix-Marseille University by Servant et al. (2014). The paradigm and methods of the study are summarized here, but the reader is encouraged to read the original paper for further details. Participants were shown arrays of circles, and were asked to respond as to whether the color of the center circle was red or blue. After providing informed consent, participants received instructions, completed a practice block, then began the task. Each trial began with the appearance of three circles, which remained on the screen until participants responded with a maximum duration of 1500 ms. After the stimulus was removed from the screen, there was an inter-trial interval of 1500 ms. Color-mappings were counterbalanced between participants, such that half of the participants were instructed to respond ‘left’ to a red target and ‘right’ to a blue target, and the other half were instructed to respond ‘right’ to a red target and ‘left’ to a blue target.

3.3.2.1.2 Stimuli and apparatus

Participants completed 24 blocks of the task, each block containing 96 trials (2,304 trials in total). Stimuli were presented using PsychoPy software (Peirce, 2007) on a CRT color monitor with a refresh rate of 100Hz. Flanker circles could be the same color (congruent) or a different color (incongruent) relative to the target. Importantly, the color saturation of center target circles varied from trial to trial within six conditions (degrees of suprathreshold saturation levels: 15%,

25%, 35%, 45%, 60% and 80%), while the color saturation of flanker circles was held constant at 80%. Task condition (congruent or incongruent), target hue (red or blue), and target color saturation (6 levels) were counterbalanced within block. Stimuli appeared along the horizontal midline of a black field. To respond, participants made left or right button presses with their corresponding thumb. Buttons were set atop plastic hand grips that were 3 cm in diameter and 7 cm in height, with 20 cm in between. Examples of the stimuli are provided in Figure 3.13, based on Figure 2 in Servant et al., 2014.

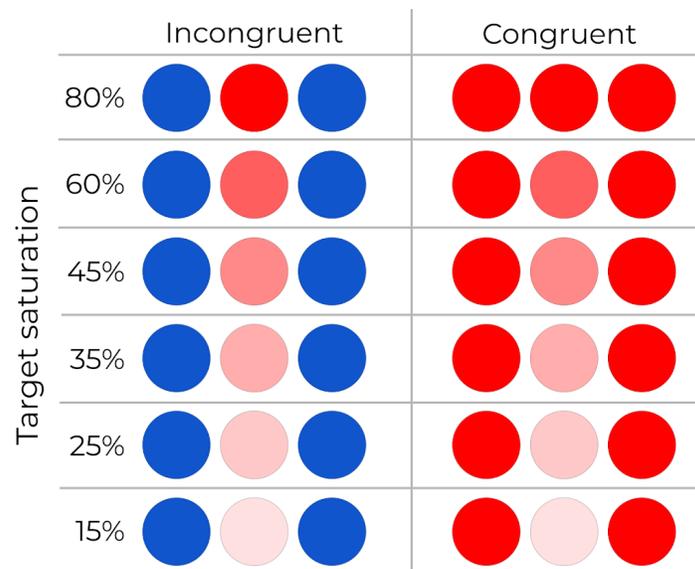


Figure 3.13: Examples of stimuli used in Experiment 2, based on Figure 2 in Servant et al., 2014. Each stimulus consisted of a target circle (red or blue), flanked by two circles of an incongruent (Left column) or congruent (Right column) color. Targets varied in saturation between 15 and 80% (rows) while the color saturation of the flankers was held constant at 80%. While only stimuli with red targets are shown here, the paradigm was counterbalanced so that 50% of stimuli featured a blue target.

3.3.2.1.3 Participants

Twelve students provided informed consent in accordance with the Declaration of Helsinki, and participated in the study in exchange for 10€/hour. Participants had normal or corrected-to-normal vision and normal color vision.

3.3.2.1.4 Model-fitting

Prior to fitting the models, we first needed to make adjustments to the models to accommodate the conditions of the target saturation manipulation. Following the example of Servant et al. (2014), we made the assumption that the p parameter, representing perceptual input strength that behaves within the SSP as a scalar on the spotlight, was the logical candidate for tracking the perceptual strength of each item in the stimulus array. We therefore modified all models of interest to include six separate values of p representing the six conditions of target saturation included in the experiment. Drift rates ρ_1 and ρ_2 for each accumulator were calculated via the following modifications to Equations 3.2 and 3.3:

$$\text{congruent} : \rho_2 = p_C a_{\text{target}} + 2p_{0.80} a_{\text{flanker}}; \rho_1 = 0$$

Equation 3.5

$$\text{incongruent} : \rho_2 = p_C a_{\text{target}}; \rho_1 = 2p_{0.80} a_{\text{flanker}}$$

Equation 3.6

where $C \in \{0.15, 0.25, 0.35, 0.45, 0.60, 0.80\}$ and was selected depending on the color saturation of the target in each trial. In Equations 3.5 and 3.6, a_{flanker} was always scaled by $p_{0.80}$ since the color saturation of flanker stimuli was held constant at 80% across trials. Values of p_C were constrained so that $p_{0.15} \leq p_{0.25} \leq p_{0.35} \leq p_{0.45} \leq p_{0.60} \leq p_{0.80}$. In each model, the

vector of values k such that $k = [p_{0.15}, p_{0.25}, p_{0.35}, p_{0.45}, p_{0.60}, p_{0.80}]$ was calculated via a sigmoidal function

$$k_i = \frac{a}{1 + e^{-c(h_i - b)}}$$

where $h = [0.15, 0.25, 0.35, 0.45, 0.60, 0.80]$ and a , b , and c were free parameters. We decided on this parameterization because we assumed perceptual input strength values of p_C varied monotonically as a function of perceptual strength, but did not have any strong hypotheses about the functional form of the relationship among them. The sigmoidal function provided an appropriate level of constraint while still being able to capture a wide variety of curves as illustrated in Figure 3.14.

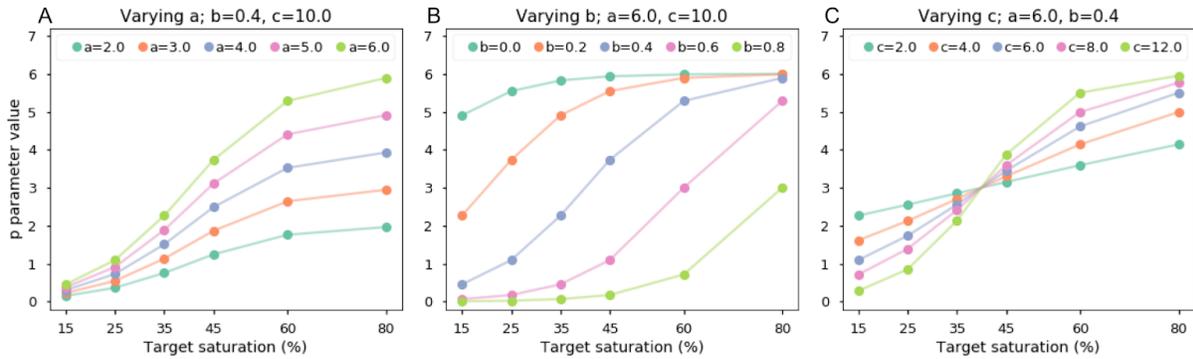


Figure 3.14: Range of sigmoid functions for calculating p_c . Sigmoid functions were implemented to capture the attention allocated to stimuli in the six saturation conditions in Experiment 2.

Panel A shows the effect of modifying the a parameter while keeping b and c constant. Panels B and C similarly show the effects of modifying the b and c parameters respectively, while the other parameters are held constant.

Priors for parameters a , b , and c were selected to be mildly informative, and were defined as follows:

$$a \sim \mathcal{TN}(1, 10, 0, 20)$$

$$b \sim \mathcal{U}(-1, 10)$$

$$c \sim \mathcal{TN}(4, 10, 0, 30)$$

Priors for all other parameters as well as all model fitting procedures were otherwise identical to those described for Experiment 1. We modified the DSTP to include a sigmoid function for fitting the target color saturation conditions as well. Details of the modified DSTP models are included in Appendix B.

3.3.2.2 Results

3.3.2.2.1 Behavior

Responses shorter than 150 ms were excluded from analyses and model-fitting (<0.01% of trials across subjects). Detailed behavioral results of Experiment 2 are presented in Servant et al. (2014). In summary, participants were significantly slower ($t(11)=6.491$, $p<0.001$) and less accurate ($t(11)=-3.437$, $p<0.01$) on incongruent trials relative to congruent, across target color saturation conditions. Participants were also significantly slower (15% saturation - 80% saturation: $t(11)=11.583$, $p<0.001$) and less accurate (15% saturation - 80% saturation: $t(11)=7.425$, $p<0.001$) on lower saturation trials relative to higher saturation trials, and the effect persisted both within incongruent (RT: $t(11)=9.109$, $p<0.001$; accuracy: $t(11)=6.390$, $p<0.001$) and congruent trials (RT: $t(11)=11.646$, $p<0.001$; accuracy: $t(11)=7.571$, $p<0.001$). Table 3.3 contains mean RTs and error rates in each condition of Experiment 2.

Table 3.3: Average mean RTs (ms) and error rates (in parentheses) across participants for Chapter 3, Experiment 2

| Condition | Target Saturation | | | | | |
|-------------|-------------------|-------------|-------------|-------------|-------------|-------------|
| | 15% | 25% | 35% | 45% | 60% | 80% |
| Incongruent | 477 (0.326) | 458 (0.224) | 443 (0.154) | 437 (0.132) | 425 (0.114) | 422 (0.107) |
| Congruent | 449 (0.142) | 421 (0.081) | 410 (0.053) | 399 (0.043) | 391 (0.041) | 386 (0.047) |

3.3.2.2.2 Model fits

BPIC values for each model were mean-centered within subject, and are shown as a heat map in Figure 3.15. The LCA control model was the winning model in 8 out of the 12 participants, the FFI control model was the winning model for 3 participants, and the DSTP was the winning model for 1 participant. Accounting for the magnitude of the wins across subjects, the LCA control model outperformed all alternatives, including FFI control, while the FFI time and DSTP models fit the worst overall.

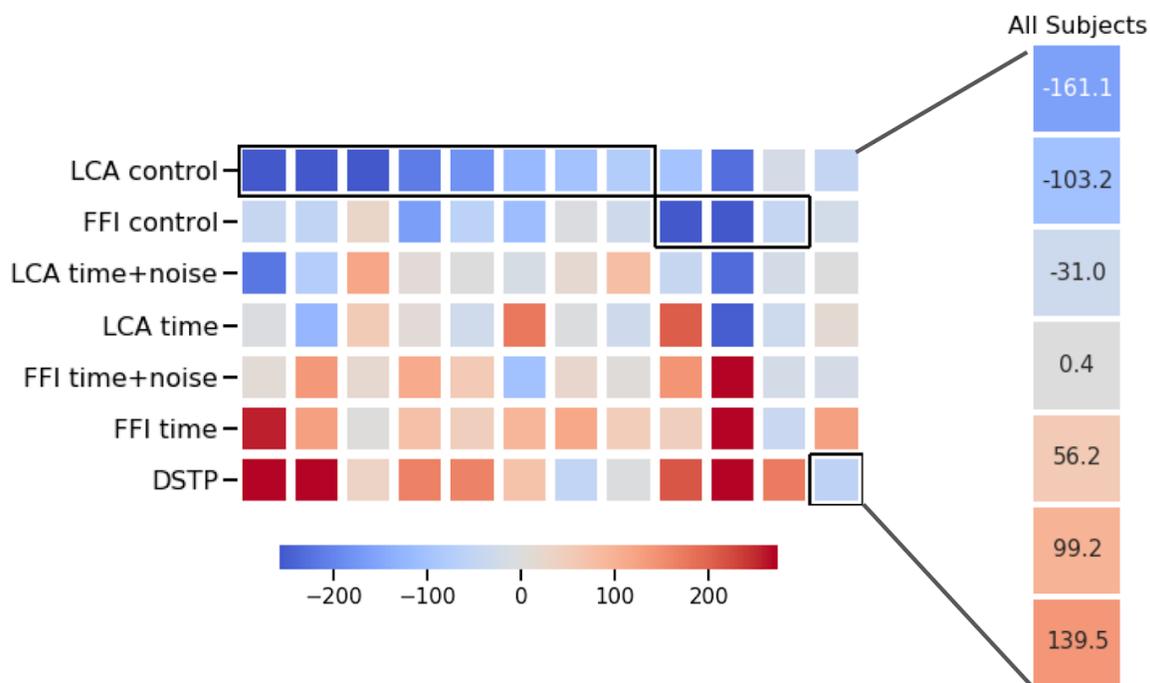


Figure 3.15: Heat map of BPIC values, mean-centered within-subject for Chapter 3, Experiment 2. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right.

Figure 3.16 includes observed choice-RT distributions for each task condition (congruent and incongruent) and target color saturation condition (low: 15%, 25%, 35% and high: 45%, 60%, 80%), averaged across participants. Mean distributions generated from each subject's best-fitting parameters in our four main models of interest are shown as well. Similar to the results of Experiment 1 shown in Figure 3.7, the two control-based models provided better qualitative fits to the RT distributions for correct responses, compared to the time-based models. This again reflects the ability of the control-driven models to capture the nuanced differences in behavior across subjects, specifically subject-level differences in fast and slow responses across conditions

due to the nature of the control signal. More importantly, Figure 3.16 shows that the FFI and LCA models make drastically different predictions about the error RT distributions, particularly in the incongruent condition. While the LCA models are generally able to capture the peak and spread of the incongruent error RTs, the FFI models consistently predict a larger proportion of fast errors across target color saturation conditions than we observe in the data. This overprediction of fast errors is a natural consequence of the strongly-correlated evidence accumulation mechanism in the FFI models. The FFI models are able to predict different drift rates across saturation conditions due to differences in the perceptual input strength scaling parameters (p_c), and are therefore able to capture the general pattern of faster correct responses for high target saturation trials. Because of the strongly-correlated evidence accumulation mechanism, however, faster positive drift rates for one accumulator result in correspondingly faster *negative* drift rates for the other. As such, the FFI models are limited in their ability to concurrently capture observed RTs for correct and error responses across all conditions. In contrast, the flexibility of the weakly-correlated evidence accumulation mechanism in the LCA models allow the models to seamlessly adapt to conditions of unequal perceptual strength between target and flanker stimuli.

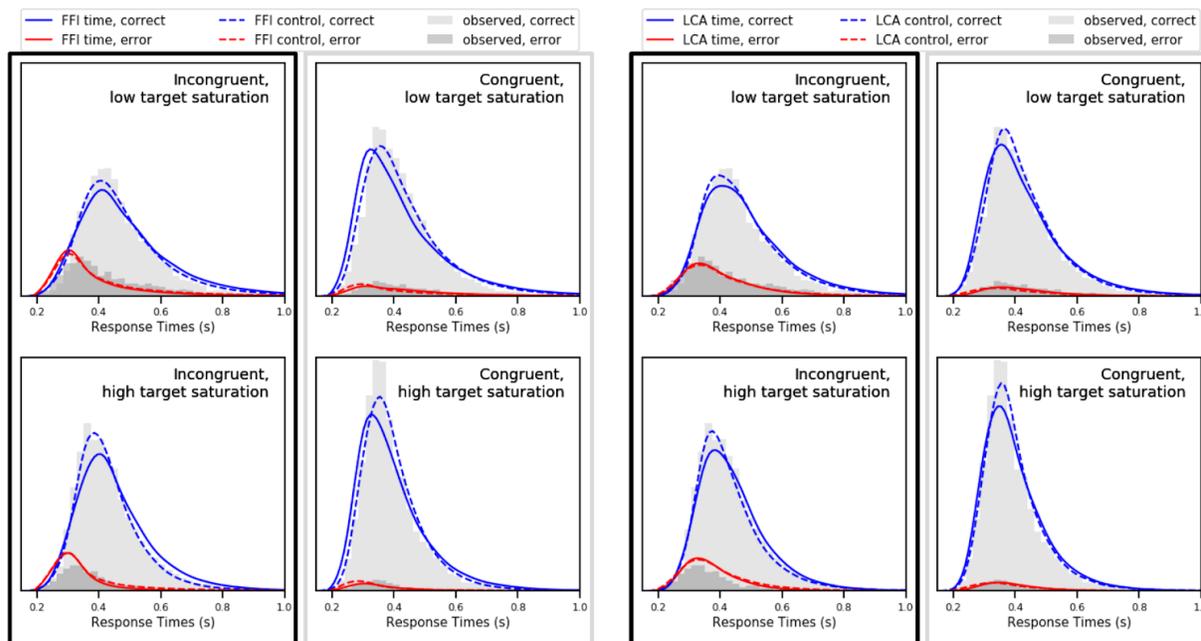


Figure 3.16: Observed and model-generated choice-RT distributions. Observed RT distributions for correct (light gray histograms) and incorrect (dark gray histograms) responses were averaged across participants. Models were simulated 10,000 times for each condition, using each participant's best-fitting parameters. Lines show average model-generated distributions across participants. Distributions generated by the FFI time and FFI control models are shown in the left panel, and distributions generated by the LCA time and LCA control models are shown in the right panel. Choice-RT distributions for low target saturation trials are shown in the top row and high saturation trials are shown in the bottom row.

3.3.2.3 Discussion

We hypothesized that the flanker saturation manipulation in Experiment 2, in which targets and flankers differed in perceptual strength from trial to trial, would cause models with strongly- and

weakly-correlated evidence accumulation mechanisms to make contrasting predictions. Because an increase in evidence for one choice option results in an equivalent decrease in evidence for the other choice, the FFI models do not predict any mechanistic differences for how participants process stimuli across different target saturation conditions. These models, therefore, depend on the values of the perceptual input strength scalars p_c to capture any behavioral differences between high- and low-saturation target conditions. As shown in Figure 3.16, however, the FFI models were only able to approximate RTs for correct responses at the expense of the error distributions--both the FFI time and the FFI control models predicted faster error RTs in the incongruent condition. The LCA models were more successful overall compared to the FFI models at fitting the shapes of all choice-RT distributions across saturation and congruency conditions, suggesting that the flexibility afforded by a weakly-correlated evidence accumulation structure is necessary for fitting these data.

Consistent with the results of Experiment 1, models with control-based attention mechanisms provided better fits to the data compared to time-based alternatives. Despite being the most complex model in our comparison with 14 free parameters (compared to 7 in FFI time, 8 in FFI time+noise and FFI control, 9 in LCA time, and 10 in LCA time+noise and LCA control), the DSTP provided the worst quantitative fits as determined by BPIC. We included the DSTP in the current project to test our control-based attention mechanism against an alternative decision-based mechanism. The results of Experiments 1 and 2 indicate that our control-based mechanism strikes a more effective balance between flexibility and parsimony than the DSTP.

Taken together, the results of Experiment 2 indicate that both LCA evidence accumulation mechanisms and control-driven attention mechanisms are necessary for appropriately predicting behavior under conditions of differing perceptual strength.

3.3.3 Experiment 3

One motivation for the current project was to develop a neurally-plausible mechanism for modulation of attention within-trial. Our theory, which we operationalized via our cognitive control-based models, is that modulation of attention is an emergent property of the dynamics of the decision process. While we do find evidence for cognitive control-based processes across Experiments 1 and 2 by fitting our models to behavioral data alone, we wished to determine whether our model-generated signal for cognitive control actually maps onto an observable, within-trial signal in the brain. In Experiment 3, we collected EEG data alongside the same standard flanker task administered in Experiment 1 and designed a model-based EEG analysis with a latent input approach to gain insight into the within-trial processes that we could not observe from behavior alone. Based on the results of Experiments 1 and 2, we predicted that LCA mechanisms in combination with control-based attentional mechanisms would most effectively track latent EEG measures.

3.3.3.1 Methods

3.3.3.1.1 Procedure and EEG Acquisition

Participants completed a standard flanker task that was identical to the one administered in Experiment 1. After providing written informed consent, participants were fitted with an elastic cap embedded with 64 Ag-AgCl active scalp electrodes arranged in an extended 10-20 array

(BrainProducts GmbH, Munich, Germany), and seated in an electrically-shielded, sound-attenuated testing room. Participants were asked to turn off all electronic devices and leave them outside of the testing room before the experiment began. The EEG signal was sampled at a rate of 1000 Hz via a DC-powered actiCHamp amplifier connected to a desktop PC. The ground electrode was located at Fpz and the reference was set to the average of mastoid electrodes TP9 and TP10 during recording. Electrode impedances were reduced to less than 25Kohms via application of electrolyte gel as recommended by the equipment manufacturer. Instructions for the task appeared on the computer screen, and were read aloud by the experimenter. Participants were given the opportunity to take breaks from the task in between task blocks, but remained seated in the testing room throughout. During each trial, a fixation cross appeared in the center of the screen for 1000 ms before being removed. The trial stimulus then appeared on the screen after a jittered duration of 100-900 ms. Participants responded by pressing the 'J' key on the keyboard if the arrow in the center of the array pointed left, and the 'K' key if the center arrow pointed right. Participants were asked to respond with their right forefinger and right middle finger respectively. Only responses made 150 ms after the stimulus appeared were recorded, and the stimulus was removed from the screen immediately after the participant made a valid response. Participants were given an unlimited amount of time to respond, but were instructed to respond as quickly and accurately as possible. EEG signal was monitored by the experimenter throughout the session for abnormalities using PyCorder software (BrainProducts GmbH, Munich, Germany) on the acquisition PC.

3.3.3.1.2 Stimuli and apparatus

Stimuli were presented and recorded via a desktop PC equipped with Linux OS connected to a 24” LCD display with a refresh rate of 120Hz. As in Experiment 1, stimuli were presented via a custom program in SMILE. Stimuli were presented in white text on the horizontal midline of a dark gray field. Arrays on each trial consisted of a central target arrow pointing left or right, accompanied by 3 flanker items to the left and right that could be congruent (same direction), incongruent (opposite direction) or neutral (lowercase ‘o’ characters) relative to the target. Participants completed 20 blocks of the task, each block containing 48 trials that were counterbalanced by condition (congruent, incongruent, neutral) and target direction (left, right). In total, each participant completed 960 trials.

3.3.3.1.3 Participants

8 right-handed participants who were fluent in English were recruited from The Ohio State University, and were compensated at a rate of \$10/hour. All participants provided informed consent in accordance with the requirements of the Institutional Review Board at the university.

3.3.3.1.4 Model fitting

Models were fit to behavioral data only, using procedures identical to those described in the methods for Experiment 1.

3.3.3.1.5 EEG preprocessing

All EEG preprocessing was completed using custom functions in the software package Python Time Series Analysis (PTSA; <https://github.com/compmem/ptsa>). Data were filtered at 30 Hz to

eliminate low-frequency noise, and were resampled to 100 Hz to match the time step parameter dt used in our model-fitting procedure. We employed wavelet-enhanced independent component analysis (wICA; Castellanos & Makarov, 2006) to remove artifacts from eye-blinks and saccades. Trials were segmented into epochs and time-locked to when the stimulus appeared on the screen. Epochs were 2500 ms long beginning 500 ms before stimulus onset, and were baseline-adjusted according to the mean voltage within a 200 ms pre-stimulus window. Epochs were rejected if kurtosis exceeded 5.0 or if the amplitude range exceeded 100V (17% of all trials).

3.3.3.1.6 Model-based EEG analysis

Given that the models in our investigation make different predictions about the behavior of the attentional spotlight within each trial, we wanted to determine which mechanism best mapped onto observed neural signals. As such, we used within-trial correlation analyses to assess the link between model-generated attention signals and EEG voltages at each electrode. Here, the “attention signal” refers to the function that controls the shrinking of the attentional spotlight in between stimulus onset and response within each model. In a time-based model, the attention signal would be a vector of t values for calculating spotlight width at each time step as described by Equation 3.1 in Section 2.2.1. In contrast, the attention signal in a control-based model would be a vector of c values representing a continuous measure of cognitive control for calculating the spotlight width as shown in Equation 3.4.

We first fit the models to behavioral data from each participant, and identified MAP estimates for the parameters. The procedures that follow are described in terms of a single model, but were

repeated for LCA control, LCA time, LCA time+noise, FFI control, FFI time, and FFI time+noise. Using each set of best-fitting parameters, we generated 30,000 trials per task condition. Each simulation produced a choice (correct or incorrect), RT, and a vector of values representing the within-trial attention signal through time. We then matched each observed response made by the participant to a subset of responses generated by the model using the same participant's best-fitting parameters. A "match" was determined based on the following criteria: 1) the task condition of the simulation was the same as the task condition of the observed trial; 2) the choice output of the simulation (correct or incorrect) was consistent with the participant's accuracy on the observed trial; 3) the RT of the simulation fell within a window from $RT_{observed} - dt$ to $RT_{observed} + dt$, where dt is the step size for time discretization in our model-fitting procedure ($dt = 0.01$; see *Sections 2.1.1* and *2.1.2*). Observed trials that matched fewer than 100 out of the 30,000 simulated trials in at least one model were discarded from further analyses (38.5% of trials). Despite excluding a large proportion of trials, 3,914 trials across participants were still included in our final analysis.

Each remaining observed trial corresponded to a matrix of attention signal values, where each row represented a simulated trial and each column represented a time step within-trial. Mean attention signal values were then calculated across rows. The purpose of this procedure was to estimate attention processes underlying the participant's observed behavior, within the contrasting theoretical frameworks of our models. The result was a model-generated vector of attention signal values for each trial, spanning the duration of the decision process and terminating at the point when the participant made a response. The next step was to assess the correspondence between model-generated attention signals and EEG voltage within-trial.

EEG data at each electrode was preprocessed and segmented into trial-level epochs as described previously. On each trial, we defined a decision-relevant time window from $\frac{\tau}{2}$ seconds after stimulus onset to $\frac{\tau}{2}$ seconds before the response, where τ was the best-fitting non decision time parameter value for the subject at hand. We then calculated the Pearson's r correlation between the vector of within-trial attention values (with shape $[1, N]$, where N was the number of time steps between the start of the decision and the response), and the matrix of within-trial EEG voltage values (with shape $[E, N]$, where $E=64$ electrodes). After repeating this procedure for every trial, the result was a matrix M of Pearson's r values, where each row represented a trial and each column represented an electrode. We then applied a Fisher's Z -transform to matrix M to satisfy the assumptions of statistical inference. One-sample t -tests were used to calculate a p -value for each column of matrix M , where the null hypothesis was that the mean trial-level Z correlation at each electrode did not differ from 0. Significance was determined via the Benjamini-Hochberg procedure for adjusting for multiple comparisons, which entails a rank-ordering of p -values at each electrode and a sliding significance criterion (Benjamini & Hochberg, 1995). This provided a single EEG topography for each model, illustrating the extent to which model-generated attention signals significantly correlated with within-trial EEG activity. Because the DSTP model does not contain a continuous within-trial mechanism for attention modulation, we fit the DSTP to the behavioral data from Experiment 3 but did not include it in the EEG analysis.

3.3.3.2 Results

3.3.3.2.1 Behavior

Responses shorter than 150 ms or longer than 2000 ms were excluded from analyses and model-fitting (<2% of trials across subjects). As in Experiment 1, neutral trials were excluded due to unforeseen perceptual pop-out effects. A summary of behavioral results is shown in Table 3.4.

We observed a similar pattern of results as in Experiment 1, specifically lower accuracy on incongruent trials compared to congruent ($t(7)=-6.652$, $p<0.001$) and slower RTs for incongruent trials compared to congruent ($t(7)=4.935$, $p<0.05$). We observed fast errors in both conditions, but the RT difference between correct and error responses was only significant among incongruent ($t(7)=-6.392$, $p<0.001$) and not among congruent trials ($t(6)=0.187$, $p=0.858$).

Table 3.4: Average accuracy and mean RTs (ms) across participants for Chapter 3, Experiment 3

| Condition | Accuracy | All RT | Correct RT | Error RT |
|-------------|----------|--------|------------|----------|
| Incongruent | 0.936 | 738 | 756 | 486 |
| Congruent | 0.990 | 552 | 553 | 520 |

3.3.3.2.2 Condition-level EEG

Stimulus-locked ERP results for correct responses in Experiment 3 replicated standard flanker effects (Kopp et al., 1996). In central-posterior electrode locations, an N2 peak occurred 340-400 ms after stimuli appeared in the incongruent but not the congruent condition. We assessed significance by means of a non-parametric permutation test with threshold-free cluster enhancement (TFCE; S. Smith & Nichols, 2009). Each participant's data were randomly shuffled

500 times with replacement, and we performed a 1-sample t-test at the level of each participant, electrode, and time point within-trial, where the null hypothesis was that there was no difference in voltage between congruent and incongruent trials. Using a critical family-wise error threshold of $p=0.05$, we identified one cluster encompassing electrodes CP1, Cz, CPz, and P1 at time points between 350 and 380 ms post-stimulus at which the voltage difference between the congruent and incongruent conditions was significant. Topographic plots and grand average ERP waveforms at CPz for the condition-level comparison are shown in Figure 3.17.

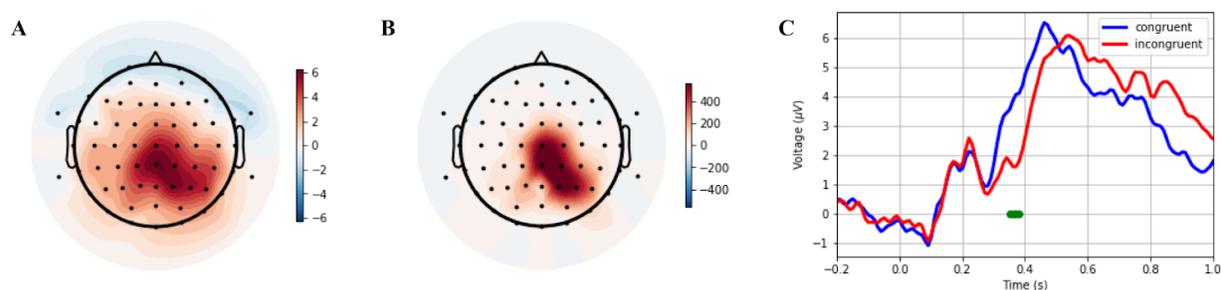


Figure 3.17: Condition-level EEG results for Experiment 3. Topographic maps show voltage differences between congruent and incongruent conditions at 370 ms post-stimulus, before (Panel A) and after (Panel B) threshold-free cluster enhancement (TFCE). Panel C shows grand average ERP stimulus-locked waveforms for congruent and incongruent trials at electrode CPz. Significant condition-level differences as determined by TFCE are shown as green points.

3.3.3.2.2 Model fits to behavior

Because we used the same task paradigm in Experiment 3 as in Experiment 1, we expected to observe the same patterns in our model fits. Indeed, goodness-of-fit as measured by BPIC values replicated the mixed results we observed in Experiment 1. When we calculate the average mean-centered BPIC values across subjects, the LCA control model outperforms the alternatives

(average mean-centered BPIC=-51.0) with the FFI control model coming in second place (average mean-centered BPIC=-28.7). A heatmap showing the full set of goodness-of-fit results is included in Appendix B.

3.3.3.2.3 Model-based EEG results

Using decision output generated from each model, we calculated correlations between the signals controlling the width of the attentional spotlight (e.g. time, time+noise, or cognitive control) and EEG voltage during the decision. Figure 3.18 illustrates the foundation of our model-based EEG analysis. Visually, we observe that the control models generate attention signals that gradually increase through time and begin to stabilize before a decision is made, similar to the EEG signals. The time and time+noise models both predict more linear signals. The time+noise models are able to predict variability in the rate of signal increase depending on the duration of the decision, but the time models predict an identical trajectory of the attention signal on every trial.

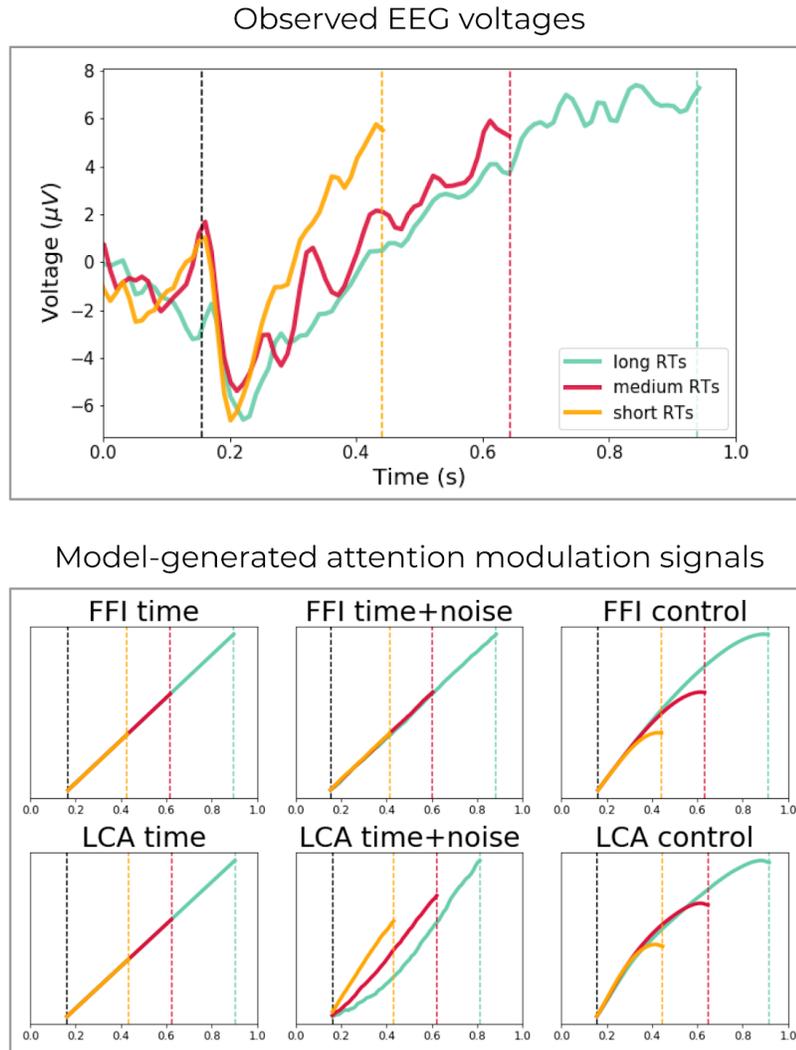


Figure 3.18: Observed EEG voltages and model-generated attention modulation signals. Data and simulations are shown for one subject. Analyses were completed at the level of every trial and electrode, but for the purposes of this visualization, EEG voltages were averaged across electrodes that demonstrated the highest correlation with model-generated attention signals (TP8, P2, C6, CP6, CPz, Pz, FC6, C2, CP1, T8, P1, P4, FC4). Data were divided into three bins based on three equal RT percentiles. Vertical lines represent the boundaries of the decision-relevant

interval between stimulus onset and the mean RT within-bin, limited by the mean best-fitting t_0 across models.

Means across trial-level Z correlations between EEG voltage and model-generated attention modulation signals at each electrode are illustrated as topographic plots in Figure 3.19. All 6 models predicted attentional mechanisms that were most correlated with EEG activity at right-posterior electrode locations. Out of all of the models, only the correlations between attentional mechanisms in the LCA control model and EEG activity were statistically significant (critical value = 0.1; electrodes TP8, P2, C6, CP6, CPz, Pz, FC6, C2, CP1, T8, P1, P4, FC4).

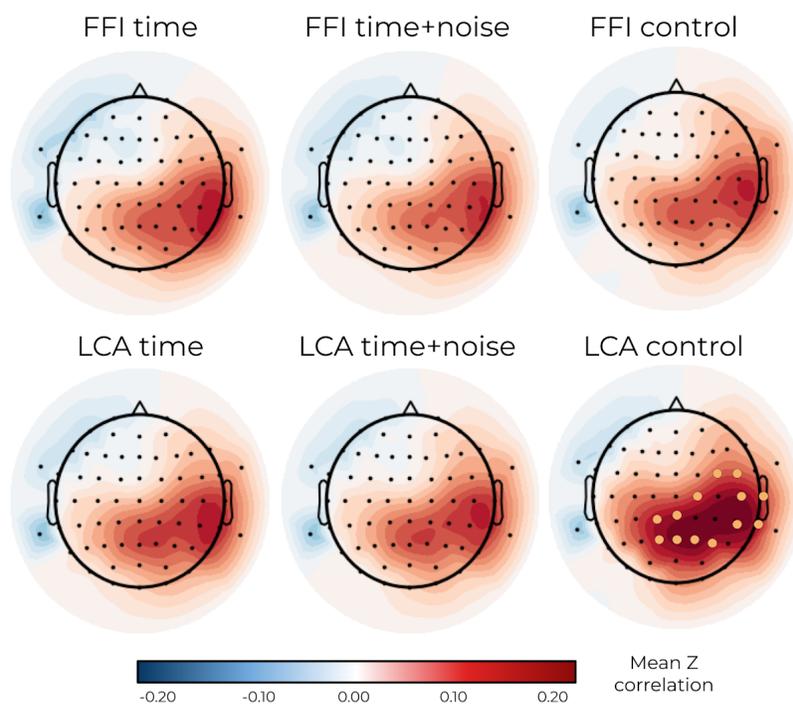


Figure 3.19: Mean Z correlation maps for observed EEG data and model-generated attention modulation signals. Data were generated by each model using each participant's best-fitting parameters. For each trial, we calculated an average vector of drive to the attention mechanism

through time using each model's simulations. Trial-level correlations between EEG voltage and model-generated attention were calculated. Pearson's r values were Fisher's Z -transformed, and p values were calculated for each model and electrode using a 1-sample t -test. Significance was determined via Benjamini-Hochberg correction for multiple comparisons, and are indicated by yellow points.

To observe differences in model predictions of attention modulation and how they relate to neural signals, we calculated the pairwise differences in model predictions and EEG voltage correlations at the level of each trial, and then calculated means at each electrode. Three comparisons yielded significant electrode-level differences: LCA control vs. FFI control (C4, C2, C1, C3, CP4, CP2, CPz, CP1, CP3, P1, Pz, P2, P4), LCA control vs. FFI time (FC2, FCz, FC1, FC3, C4, C2, Cz, C1, C3, C5, CP4, CP2, CPz, CP1, CP3, CP5, P3, P1, Pz, P2, P4, POz, PO3, Oz) and LCA control vs. FFI time+noise (FC4, FC2, FCz, FC1, FC3, C6, C4, C2, Cz, C1, C3, C5, CP6, CP4, CP2, CPz, CP1, CP3, CP5, P6, P5, P3, P1, Pz, P2, P4, PO4, POz, PO3, Oz). Topographic plots in Figure 3.20 show that increased correlations between EEG voltage and attention modulation in LCA control, relative to the predictions of the other models, are widespread across the scalp. All other pairwise difference maps are shown in Appendix B.

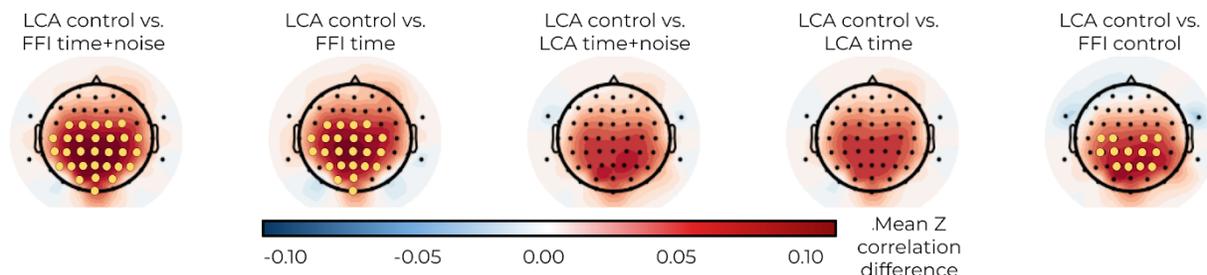


Figure 3.20: Mean Z correlation difference maps for observed EEG data and model-generated attention modulation signals. After calculating Z correlation values for each model and each electrode, we calculated the pairwise difference topographic maps for each possible pair of models. P values were calculated for each model comparison and electrode using a 1-sample t-test. Significant correlation differences were identified using a Bejamini-Hochberg correction for multiple comparisons, indicated by yellow points.

3.3.3.3 Discussion

Because we were interested in developing a neurally plausible model of the flanker task, we wanted to test whether the attention mechanisms in any of our models resembled the fluctuations of within-trial neural signals as measured by EEG. Attention mechanisms in all models were most correlated with EEG activity in right-posterior regions, as shown in Figure 3.19, but only the LCA control model yielded significant correlation results. This is an interesting pattern of findings in light of previous EEG studies designed to probe the spotlight view of spatial attention, which often reported attentional correlates at posterior electrodes as well (Awh et al., 2000; Busch & VanRullen, 2010; Handy et al., 2001). These studies, however, tended to observe attention-related activity at central-posterior electrodes, and lateralized effects only occurred

when stimuli appeared in the edges of the visual field (Hillyard et al., 1998; Mangun & Hillyard, 1988; M. Müller et al., 2003). For example, Mangun and Hillyard (1988) investigated the hypothesis that early sensory-evoked peaks would reflect a spotlight-like filtering of information. The authors identified gradual decreases in P1 and N1 amplitudes that varied as a function of distance between attended and evoking stimuli. These effects were specifically observed in posterior electrode locations, contralateral to the screen location of the attended stimuli. Because stimuli were only presented in the center of the screen in our paradigm, we believed our right-lateralized results could reflect contamination by motor effects given that participants made all responses with the right hand. Because this would result in strong motor-related activity in the left hemisphere, it potentially obfuscated the attention-related activation. It is nevertheless notable that only the LCA control model generates a within-trial attention modulation signal that significantly correlates with the gradual ramp-up and relaxation of neural amplitudes as shown in Figure 3.18 at attention-relevant locations on the scalp.

We calculated the pairwise differences maps shown in Figure 3.20 for two purposes: 1) to cancel out the motor effects that could have affected each individual model-based EEG analysis, and 2) to observe how each model compared to the others in terms of generating a neurally plausible attention modulation signal. Specifically for comparisons involving the LCA control model, we identified large differences in correlation means that were widespread across the scalp. This implies that the LCA control model was able to generate within-trial signals that resemble the general time course of EEG voltages better than the alternative models. While we do not make any strong claims here about the LCA control model capturing any specific neural processes, the

results of Experiment 3 support the notion that the mechanisms in the LCA control model behave in a way that is in line with observed voltage time courses in the brain.

3.4 General discussion

3.4.1 Summary

In the current project, we presented a mechanistic theory of cognitive control in which within-trial modulation of attention is a byproduct of interacting decision processes. We tested our theory by developing a set of SSMs, each making alternative assumptions about evidence accumulation and attention modulation mechanisms. Models included time-based attention processes like the existing flanker SSMs, or control-based attention mechanisms inspired by connectionist models (i.e. Botvinick et al., 2004; De Pisapia & Braver, 2006; Verguts 2017). Because the control-based models calculate attention modulation from the noisy accumulators while the time-based models operate in a strictly linear manner, we also included model variants that calculate attention based on time with additional random noise. When specifying the evidence accumulation processes in our models, we developed models with either strongly-correlated accumulators defined by FFI mechanisms, or weakly-correlated accumulators defined by LCA mechanisms. These two mechanisms represent different hypotheses about the neural underpinnings of the decision process: the former assumes decisions are based on the difference in firing across populations of neurons, and the latter assumes decisions are based on the competition between the two most active populations of neurons. Though the competing hypotheses concerning attention modulation and evidence accumulation were implemented and compared within the SSP model, we fit the DSTP model as an additional point of comparison as well. The DSTP presents an alternative mechanistic explanation for decision-guided attention, in

which response selection processes are conditionally dependent upon the outcome of stimulus selection processes. Across three experiments, we found evidence that weakly-correlated LCA mechanisms in combination with dynamic, control-guided attention modulation mechanisms best-accounted for the data in each task condition.

In Experiment 1, we fit the models to data from a standard flanker task. While all models fit the data well, the two control-based models provided the best fits as determined by BPIC. Further insights from ELI and CAF analyses revealed that the LCA control model was particularly effective at capturing nuanced differences in performance between subjects, including slow errors in the incongruent condition and fast errors in the congruent condition. To hone in on the mechanistic assumptions of the FFI and LCA mechanisms, Experiment 2 featured a manipulation of target color saturation. Because the FFI models assume that an increase in evidence for one response requires a decrease in evidence for the other, we found that the FFI models overestimated the speed of error distributions across conditions. The LCA models, and particularly the LCA control model, were more flexible and therefore able to capture behavior under conditions where targets and flankers differed in perceptual strength. In Experiment 3, we collected EEG data alongside a standard flanker task in an effort to determine if any of our model-generated attention modulation signals resembled within-trial processes in the brain. Using a model-based EEG analysis with a latent input approach, we found that the within-trial control signal generated by the LCA control model uniquely mapped onto the time course of EEG voltages in between stimulus onset and response. In an effort to summarize fit results across experiments, Figure 3.21 illustrates across-subject rank order sums, normed within experiment

such that lower values indicate better fits. Considering our results together, the LCA control model was the best-fitting model compared to all other alternatives.

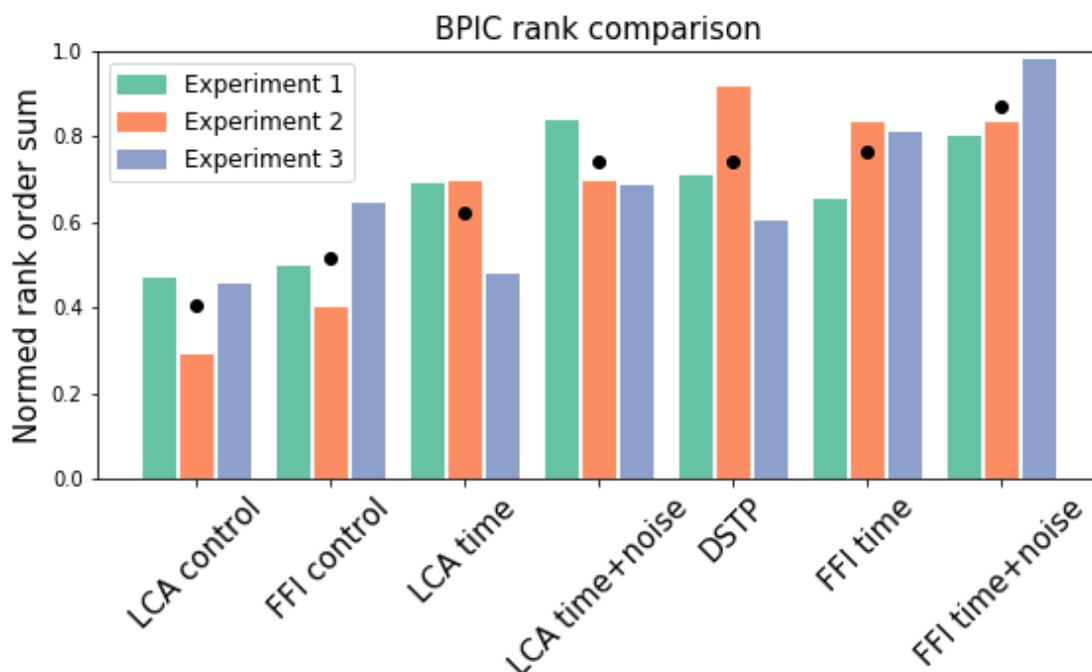


Figure 3.21: Rank order sums of BPIC values for each model and experiment. The best-fitting model for each subject and experiment as determined by BPIC was assigned a rank of ‘1’, the second best model was ranked ‘2,’ and so on. Rank values were summed within-experiment and normed based on the number of subjects in each experiment. Black points indicate mean normed rank order sums across experiments

3.4.2 Interpretation of results

In the current project, we aimed to address a gap in the literature concerning within-trial mobilization of cognitive control and modulation of attention. Several dominant theories suggest that cognitive control operates on multiple timescales to appropriately focus attention on goal-relevant information while also conserving cognitive resources (Braver et al., 2008; e.g. J. Brown

et al., 2007; Davelaar, 2008). These theories have often been operationalized within connectionist models, which feature biologically-inspired mechanisms for engaging cognitive control as a direct response to mutual activation of multiple choice units. Connectionist models, however, typically include within-trial mechanisms only en route to explaining between-trial effects, such as improved accuracy on flanker trials immediately following errors. Theories specifically designed to explain trial-level effects, such as fast errors in the incongruent flanker task condition, have instead been implemented within the SSM framework as variants of the single-accumulator DDM (Hübner et al., 2010; Ulrich et al., 2015; White et al., 2011). These models make specific predictions about attention processes that vary as a function of time, and mutually-inhibitory evidence accumulation mechanisms. Here, we introduced an SSM in which modulation of attention via cognitive control occurs as an emergent property of the dynamics of the decision process. Our model draws upon neurally-plausible mechanisms from connectionist implementations such as continuously-updated cognitive control and flexible evidence accumulation mechanisms, but was implemented in an SSM framework to allow for trial-level data-fitting and quantified model comparisons.

Despite being designed to fit data from tasks that present conflicting information, the existing flanker SSMs do not include mechanisms for tracking or modulating parameters based on mutual activation of two options. Changes to drift rate occur as a function of time, regardless of the state of competition between the two choice alternatives. By considering only the difference in activation of the two choices, these models are potentially missing an important piece of the story concerning how the brain recruits cognitive control. Furthermore, the single-accumulator structure of the flanker SSMs make the powerful assumption that an increase in evidence for one

choice results in a decrease in evidence for the other. Given the assertion that inhibitory control decisions involve two separate routes of processing, automatic and controlled, it may be overly constraining to assume that evidence accumulation between two choices is perfectly anticorrelated. By developing separate groups of models with strongly-correlated FFI mechanisms and weakly-correlated LCA mechanisms, we aimed to directly test and compare competing hypotheses about how the brain represents competing information in inhibitory control tasks. While both FFI and LCA models were able to capture general behavior in a standard flanker task as shown by the results of Experiment 1, LCA processes were important for capturing subject-level differences in performance. The perceptual strength manipulation in Experiment 2 further dissociated the predictions of the FFI and LCA models. Models with FFI mechanisms failed to appropriately capture error distributions for incongruent trials across target saturation conditions, while the flexibility of the LCA models resulted in more successful fits. Together, these findings may suggest that decisions on inhibitory control tasks may be based on the direct competition between choice options as represented by weakly-correlated mechanisms in the LCA model, rather than the difference between them. Our results seem to stand in contrast to recent findings from a stop-signal study, which found that perfect negative dependence between racing accumulators predicted aspects of observed behavior better than independent accumulators (Colonius & Diederich, 2018). This, perhaps, is indicative of mechanistic differences between 2-alternative choices and go-nogo choices, or indicates that accumulator dependence exists as a gradient and manifests differently from task to task as has been suggested in the past (P. Smith & Ratcliff, 2004). Because it has been shown that the LCA model can mimic a standard DDM under conditions of balanced leak and lateral inhibition (Bogacz et al.,

2006), the most parsimonious assumption favors the model that is flexible enough to capture all observed patterns in the data.

We hypothesized that within-trial attentional mechanisms were based on some element of the decision process rather than the mere passage of time. As such, we defined sets of models with attention mechanisms driven by time like the original SSP, models driven by time with added variability, and models driven by cognitive control which was calculated from the accumulators at each timestep within the decision process. In Experiment 1 and even more strikingly in Experiment 2, the control models outperformed the time-based models in terms of fits to behavioral data. It is important to note that the control models consistently fit the data better than time models with added variability, indicating that control mechanisms were tapping into a signal present in the data beyond random noise. In Experiment 3, this contention was reinforced by model-based EEG findings, indicating that the LCA conflict model was the only one with a time course of visual attention mechanisms that significantly correlated with within-trial EEG voltage.

Our findings provide a model-based, mechanistic complement to recent neuroimaging work that has investigated attention processes within-trial. One study recorded EEG data while participants completed a variant of the flanker task with a manipulation of visual probe locations. Probes were presented at different distances from the target on each trial in order to force modulation of the visual field (Nigbur et al., 2015). N1 ERP amplitudes, which have been shown to be an index of spatial attention (H. Heinze et al., 1994; Mangun & Hillyard, 1988), provided evidence that conflict resolution on incongruent trials occurred mainly via target enhancement, not distractor

suppression. The critical difference between Nigbur et al.'s findings and our own is that the N1 ERP reflects early perceptual processing 150-200 ms after stimulus onset (Haider et al., 1964), which is distinct from decision-related processes of interest in the current study. Considering the two sets of results together, it is possible that initial stimulus-processing in the spotlight framework of attention depends on target enhancement only, but that higher-order decision processes require additional distractor suppression mechanisms. Indeed, previous studies in EEG (Philiastides et al., 2006; VanRullen & Thorpe, 2001) have shown that visual processing and decision-making reflect distinctly different mechanisms. Philiastides and colleagues (2006), for example, recorded EEG data while participants indicated either the color or category of stimuli with different levels of phase coherence. The researchers showed that a negative ERP at 170 ms post-stimulus onset reflected identification of the goal-relevant feature in a trial (color vs. category), and that later ERPs reflected components of the decision process (red vs. green or face vs. car). Importantly, only the late ERP components reflected trial-level difficulty or conflict between the two competing choice options. Nevertheless, further work is needed to understand the possible dissociation between perceptual processing and decision-relevant computations in the presence of conflict.

Despite converging findings across three experiments, the current study is not without limitations. First, we mathematically defined within-trial cognitive control as the cumulative distance between total evidence and a learned threshold at which conflict can be resolved. We defined this function based on the DMC framework of Braver and colleagues (Braver, 2012; De Pisapia & Braver, 2006), in which cognitive control increases within-trial until attention is sufficient to resolve conflict, and then may decrease toward the end of a trial if no further

recruitment of attention is needed. Both increasing and decreasing properties were observed in neuronal firing patterns in the conflict-relevant ACC during a recent single-unit recording study (Hunt et al., 2018). Our specific definition of the cognitive control function, however, may not be precisely correct in terms of representing drive to attentional mechanisms. For example, a related mechanism described by Yeung and colleagues (2004) calculated conflict as the product of activations across possible responses. Within the SSM framework, however, the product of activations would result in an unchanging attentional spotlight if one accumulator sporadically reached zero, which would be a frequent occurrence on congruent trials. While it seems possible that the attentional spotlight would not be necessary on congruent trials, Servant and colleagues (2014) compared the original SSP to a variant in which the spotlight only shrank on incongruent trials. The authors found that the alternative model provided worse fits to behavioral data compared to the original model, and was specifically unable to capture the range in performance across subjects in the congruent condition. Future work will investigate the nature of the cognitive control signal as it relates to the amount of evidence in the system at a given time.

A second limitation is that we investigated competing hypotheses within the SSP model. We made this choice despite results from other studies demonstrating that the SSP cannot capture patterns of data beyond the flanker task (notably, negative-going delta functions in the Simon task; Ulrich et al., 2015), and that a version of the SSP implemented in the LCA framework could not capture pre-motor partial error responses as measured by MEG (Servant et al., 2015). We believe with modifications such as those explored in the current project, the shrinking spotlight framework can indeed extend beyond what it was designed to capture. Preliminary

investigations of extensions for the LCA control SSP model presented here are currently underway, specifically for tasks involving gradations of conflict outside of the flanker paradigm.

3.4.3 Conclusions

In the current study, we sought to investigate the possibility of within-trial modulation of attention based on the dynamics of the decision process, within a modeling framework that is amenable to quantifiable comparisons. We systematically developed and compared models that featured time-based or control-based attention mechanisms, and strongly- or weakly-correlated evidence accumulation mechanisms. Across three experiments, we found that a flexible accumulator structure in combination with control-based attention processes provided the best fits to behavioral data. Additionally, we found that the within-trial attention modulation signal in the LCA control model uniquely correlated with neural signals in the brain. While we have focused on within-trial mechanisms in the current study, future work will investigate the possibility that the decision-related signals driving the within-trial effects of interest here can also result in between-trial effects, such that the end-state of cognitive control in one trial contributes to the starting point of the attentional spotlight on the next.

Chapter 4: Individual Differences in Attention Allocation during a 2-Dimensional Inhibitory Control Task

In Chapter 3, I presented a set of models that were intended to mathematically formalize competing hypotheses for dynamic attention and cognitive control processes. With the exception of the DSTP, all of the models contained elements of the SSP developed by White, Ratcliff, and Starns (2011) to represent visual attention processes. While the SSP conceptualized attention as a 1-dimensional Gaussian distribution, the verbal theory on which the model was based specified a 2-dimensional spotlight that processes information in both the horizontal and vertical spatial directions. In Chapter 4, I extended the best-fitting model from Chapter 3 for processing 2-dimensional stimuli. I then capitalized on my model based approach to investigate individual differences in the natural shape that the attentional spotlight takes on at the start of each trial. Previous group-level behavioral analyses have suggested that the spotlight takes the shape of a horizontally-biased ellipse, but a model-based approach shows that this is not necessarily the case for all subjects. As mentioned previously, the manuscript presented in Chapter 4 is under review at *Attention, Perception, & Psychophysics*.

4.1 Introduction

Although the mechanisms underlying visual attention remain a topic of active research, it is widely accepted that the focal area of attention can vary in size (Castiello & Umiltà, 1992; C. Eriksen & St James, 1986; Henderson, 1991; Jonides, 1983; LaBerge, 1983; Lavie, 1995) and that processing is more efficient when the focal area is small, compared to when it is large (C. Eriksen & Schultz, 1979; C. Eriksen & Yeh, 1985; Pan & Eriksen, 1993; Umiltà, 1998). The Eriksen flanker task (B. Eriksen & Eriksen, 1974), which requires participants to respond to a target while ignoring target-congruent or -incongruent distractor stimuli, has proven to be a particularly useful tool for studying attention. Notably, Gratton and colleagues (1988) used the flanker task to reveal a time-related component to attentional processing efficiency, such that incongruent distractors cause markedly less processing interference as RTs increase. These and other behavioral and electrophysiological findings suggest that distractor stimuli have an influence on processing early in the trial, but that it decreases through time (Burle et al., 2002; Czernochowski, 2015; Nigbur et al., 2015; Ridderinkhof, 2002). In their seminal theory, Eriksen and St. James (1986) provided the explanation that attention behaves as a *zoom lens* or *shrinking spotlight* that starts out wide and diffuse at the beginning of a trial and gradually focuses on the target.

The size and shape of the attentional spotlight has been extensively studied using mixtures of horizontally- and vertically-arranged flanker stimuli (C.-C. Chen & Tyler, 2002; Cohen & Shoup, 1993; Livne & Sagi, 2011; Vejnović & Zdravković, 2015) and visual search paradigms (Hüttermann et al., 2014; Luck et al., 1989; Panagopoulos et al., 2004). By manipulating the spatial distance, position, and stimulus onset asynchrony of distractors relative to targets, for

example, Pan and Eriksen (1993) concluded that the dimensions of the spotlight dynamically adjust from trial to trial based on the spatial configuration of the stimulus at hand. In line with these results, subsequent work showed that the spotlight can take on the shape of a ring (M. Müller & Hübner, 2002), a Mexican hat (N. Müller et al., 2005) or can be divided among non-contiguous locations (Dubois et al., 2009; M. Müller et al., 2003; Treue & Martinez-Trujillo, 2012) depending on the spatial arrangement of the stimuli and the demands of the task. Through group-level analyses of speed and accuracy, however, other studies have concluded that there is a dimensional bias to the spotlight, such that it is elliptical in shape, broadly distributed along the horizontal plane and narrowly distributed along the vertical plane (Andersen & Kramer, 1993; Feng et al., 2007; Hüttermann et al., 2013). While we acknowledge that the notions of a stimulus-dependent spotlight and a horizontal attention bias are not mutually exclusive, we contend that the extent to which these processing features tradeoff within individual participants has not been thoroughly investigated.

In the current study, we used a 2-dimensional flanker task paradigm and a corresponding computational model to investigate the hypothesis that individuals vary in dimensional biases related to attentional allocation when controlling for the spatial configuration of stimuli across conditions. As shown in Figure 4.1, stimuli in the current study were designed to be identical in spatial distribution across conditions in an effort to limit stimulus-dependent modulation of the attentional spotlight. Within each condition, we manipulated the arrangement of target-congruent and -incongruent distractors to assess behavioral biases. We fit two variants of an SSM of within-trial decision processing during the flanker task to each participant's data, which allowed us to calculate parameter estimates based on trial-level choices and RTs (Weichart, Turner, et al.,

2020). Both model variants contain an attentional spotlight, implemented as the density function for a bivariate normal distribution that narrows onto the target throughout the decision process. The shape of the spotlight is specified by separate horizontal and vertical standard deviation (SD) parameters. In the *circular spotlight* model, the shape parameters were constrained to be equal in order to reflect the horizontally- and vertically-symmetric spatial configuration of the stimuli. The alternative *elliptical spotlight* variant subsumes the circular spotlight model, and allows the horizontal and vertical shape parameters to take on different values to optimally fit each participant's data, if needed. Our results show that an elliptical rather than a circular spotlight is favored for most subjects, and demonstrate notable variability between subjects in terms of horizontal or vertical biases.

4.2 Materials and Methods

4.2.1 Participants

26 undergraduate students were recruited from the University of Virginia to participate in exchange for partial course credit. All participants provided informed consent in accordance with the requirements of the Institutional Review Board at the university.

4.2.2 Stimuli and Apparatus

A custom program using the State Machine Interface Library for Experiments (SMILE; <https://github.com/compmem/smile>) was written to present stimuli, track timing, and log responses. The experiment was administered on a desktop computer with Windows 10, connected to a 24-inch, 1920 x 1080-pixel LED display with a refresh rate of 120 Hz. Stimuli were presented in white text on a dark gray background. Participants made responses using the

outer two keys of a four-key Black Box ToolKit response pad. Stimulus arrays were comprised of 13 left- or right-facing arrows arranged in a diamond formation. Distractor arrows took on one of five configurations, as illustrated in Figure 4.1, and participants were asked to indicate the direction of the arrow in the center of the array while ignoring all distractors. The stimuli were designed in consideration of research demonstrating that distractor interference is positively correlated with proximity of distractors to target (Pan & Eriksen, 1993; Andersen & Kramer, 1993; Feng, Jiang, & He, 2007). In the *easy* condition, all distractors were congruent to the target. In the *moderate* and *hard* conditions, distractors were incongruent to the target in the outer and inner layers of the array, respectively. *Horizontal* and *vertical* conditions were included to test for asymmetries in dimension-level response competition. On each trial, the array was presented in one of eight locations around the screen. Possible locations were equidistant from the center of the screen in increments of 45 degrees. Task condition (easy, moderate, hard, horizontal, or vertical), target direction (left or right) and screen location (0, 45, 90, 135, 180, 225, 270, or 315 degrees) were counterbalanced and randomized within-block.

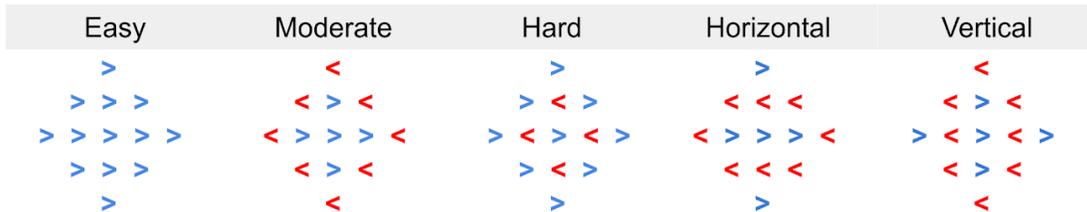


Figure 4.1: Examples of stimuli. Colors are used to highlight the contrasting directions configurations of left (red) and right (blue) arrows. All stimuli shown here contain a left-facing target, but analogous stimuli with right-facing targets were included in the experiment as well.

4.2.3 Procedure

Participants provided informed consent and were seated in individual testing rooms. Instructions and example stimuli appeared on the screen, and instructions were also provided verbally by an experimenter. A practice module with visual feedback for correct and incorrect responses was administered until the experimenter verified that the participant understood the task (~1 minute). Prior to beginning the main task, the experimenter provided the following information: “You will complete eight blocks of the task, each lasting about 2 minutes. At the end of each block, you will receive a score based on speed and accuracy. Please try to get the highest score that you can.” Once the task began, a fixation cross appeared in the center of the screen and remained for the duration of the block. Stimuli appeared on the screen until a response was made, or for a maximum of 3000 ms. Participants responded by pressing the leftmost key on the response pad if the arrow in the center of the array pointed left, and the rightmost key if the center arrow pointed right. Only responses made at least 150 ms after the stimulus appeared were recorded. At the end of each block, participants received a numerical score between 0 and 100 that was calculated as shown in Equations 4.1-4.3:

$$accuracy = \frac{\frac{N_{correct}}{N_{total}} - 0.5}{0.5}$$

Equation 4.1

$$speed = \frac{\sum_{i \in I} \frac{\log(RT_{max} + 1.0) - \log(i + 1.0)}{\log(RT_{max} + 1.0) - \log(RT_{min} + 1.0)}}{N_{total}}$$

Equation 4.2

$$score = speed * accuracy * 100$$

Equation 4.3

where I is a vector of RTs in seconds. Within this scoring metric, performance across conditions was scaled between chance (0.5) and perfect accuracy (1.0), and RTs were scaled to fall within an expected range of $RT_{min} = 350ms$ to $RT_{max} = 1350ms$. Log transforms were used in Equation 4.2 to correct for rightward skew in the RT distributions, and 1.0 was added to prevent against negative log RT values. To earn a high score, participants needed to respond both quickly and accurately in all conditions. Across 8 blocks each consisting of 80 trials, participants completed a total of 640 trials.

4.2.4 Computational Models

The base model in our current investigation was designed after the zoom lens theory of Eriksen and St. James (1986), with decision and attention mechanisms implemented within an LCA (Usher & McClelland, 2001) model framework. Specifically, we modified the *LCA-control* model of the flanker task described by Weichart and colleagues (2020) to accommodate 2-dimensional stimuli. Details of the LCA-control model are provided in the original article, but will be summarized here. In LCA-control and other accumulator models, trial-level decisions are thought to result from the noisy build-up of evidence for competing response options up to a threshold (α). Evidence accumulation is governed by drift rates that reflect the strength of information provided by the stimulus, lateral inhibition (β), and passive decay through time (κ). Each accumulator i with drift rate ρ_i and current evidence x_i is updated continuously as shown in Equation 4.4.

$$dx_i = (\rho_i - \kappa x_i - \beta \sum_{j \neq i} x_j) \frac{dt}{\Delta t} + \xi \sqrt{\frac{dt}{\Delta t}}$$

$$x_i \rightarrow \max(x_i, 0)$$

Equation 4.4

Time was discretized via the Euler method, using a step size of $dt = 0.01$ modified by a time constant of $\Delta t = 0.1$. The degree of noise is represented as a driftless Wiener process distributed as $\xi \sim \mathcal{N}(0, 1)$. Responses are made in favor of the first accumulator to exceed α , and RTs are equal to the duration of the decision process plus a nondecision time parameter (τ). Similar to the *shrinking spotlight model* designed by White, Ratcliff, and Starns (SSP; 2011), LCA-control features an attentional spotlight that gradually focuses on the target throughout the trial, but, unlike the original SSP, the spotlight shrinks due to the dynamics of the decision process and not simply time. For our current purposes, the spotlight takes the form of a density function for a bivariate normal distribution centered on the target stimulus with initial horizontal and vertical SDs of $sd_0(h)$ and $sd_0(v)$, respectively. The spotlight shrinks as a function of an online measure of cognitive control (c), modified by a rate parameter (r_d) as shown in Equations 4.5 and 4.6.

$$sd_a(h) = sd_0(h) - r_d c$$

Equation 4.5

$$sd_a(v) = sd_0(v) - r_d c$$

Equation 4.6

A ratio parameter Θ governs the relationship between $sd_0(h)$ and $sd_0(v)$, such that $\Theta = \frac{sd_0(v)}{sd_0(h)}$

. Θ was fixed to 1.0 when fitting the circular spotlight model, but was a free parameter when

fitting the elliptical spotlight model. The behaviors of the attentional spotlights in the circular and elliptical spotlight models are illustrated in Figure 4.2.

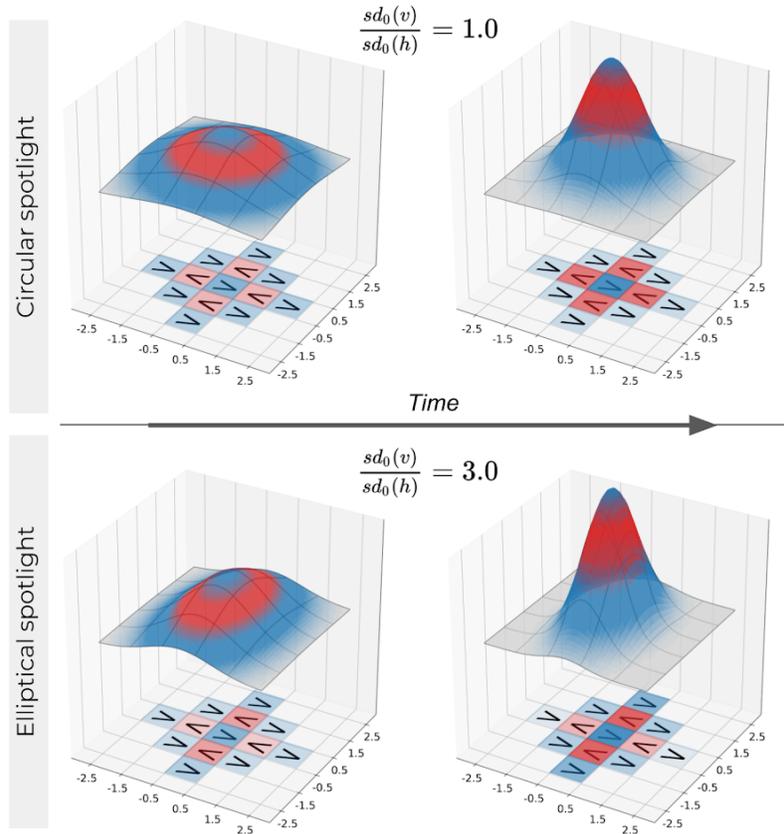


Figure 4.2: Representation of the shrinking 2-dimensional spotlight of visual attention. Over the course of a trial, illustrated from left to right, the spotlight shrinks and focuses on the center arrow of a stimulus array, rendered in each subplot at $z=0$. The strength of visual attention allocated to each arrow in a stimulus array is calculated from the density of a bivariate normal distribution within the corresponding unit square. The top panels show a circular spotlight, such that the shape parameters of the bivariate normal are constrained to be equal. The bottom panels show an elliptical spotlight, such that the two shape parameters are free to vary.

The mechanism for cognitive control is based on descriptions of reactive control discussed in Braver's dual mechanisms of control framework (Braver, 2012; Braver et al., 2008; De Pisapia & Braver, 2006), and is calculated as the cumulative distance between total evidence and a conflict resolution threshold, δ . The continuous change in c is given by Equation 4.7.

$$dc = \left(\delta - \sum_{i \in \{1,2\}} x_i \right) \frac{dt}{\Delta t}$$

Equation 4.7

Each individual arrow in a stimulus array occupies one square unit of perceptual space, and the spotlight is centered on the target. Drift rates corresponding to correct (ρ_1) and incorrect (ρ_2) responses are calculated as the total volume of the spotlight allocated to target-congruent and target-incongruent arrows, respectively. Within each unit square of the 5-by-5 stimulus array, we used the trapezoidal method to estimate the bivariate probability density at 100 equally-spaced points (Kalambet et al., 2018). The spotlight volume allocated to each unit square was then estimated from the integral of the density values over the range of interest, multiplied by the perceptual strength of a single arrow (p) as shown in Equation 4.8.

$$V = p \int_n^{n+1} \int_m^{m+1} \frac{1}{2\pi sd_a(h)sd_a(v)} \exp\left(-\frac{1}{2} \left[\frac{x^2}{sd_a(h)^2} + \frac{y^2}{sd_a(v)^2} \right]\right) dydx$$

Equation 4.8

Values of dx and dy were set to 0.1, and (x, y) coordinates fell within the unit square occupied by an arrow with vertices $(n, m), (n + 1, m), (n + 1, m + 1), (n, m + 1)$, such that $n, m \in [-2.5, -1.5, -0.5, 0.5, 1.5]$. The conceptualization of arrow stimuli within a coordinate plane is illustrated in Figure 4.2. A summary of free parameters in our two models and their respective prior distributions are shown in Table 4.1.

Table 4.1: Model parameters and priors in Chapter 4

| Parameter Description | Prior |
|-----------------------------------|--|
| Decision threshold | $\text{logit}\left(\frac{\alpha}{30.0}\right) \sim \mathcal{N}(\mu = -1.0, \sigma = 1.2)$ |
| Lateral inhibition | $\text{logit}(\beta) \sim \mathcal{N}(\mu = 0.0, \sigma = 1.4)$ |
| Evidence leak | $\text{logit}(\kappa) \sim \mathcal{N}(\mu = 0.0, \sigma = 1.4)$ |
| Nondecision time | $\text{logit}\left(\frac{\tau}{RT_{min}}\right) \sim \mathcal{N}(\mu = -0.2, \sigma = 1.2)$ |
| Initial spotlight SD (horizontal) | $\text{logit}\left(\frac{sd_0(h) - 0.1}{30.0 - 0.1}\right) \sim \mathcal{N}(\mu = -0.2, \sigma = 1.2)$ |
| Spotlight dimension ratio | $\Theta \sim \gamma(\alpha = 3.0, \beta = 2.0)$ |
| Rate of focus | $\text{logit}\left(\frac{r_d}{20.0}\right) \sim \mathcal{N}(\mu = -2.0, \sigma = 1.0)$ |
| Conflict resolution threshold | $\text{logit}\left(\frac{\delta}{30.0}\right) \sim \mathcal{N}(\mu = -1.0, \sigma = 1.2)$ |
| Perceptual input strength | $\text{logit}\left(\frac{p}{20.0}\right) \sim \mathcal{N}(\mu = -0.8, \sigma = 1.2)$ |

4.2.5 Model Fitting and Assessment

We used the model-fitting procedures described in detail by Weichart and colleagues (2020) with probability density approximation methods (PDA; Turner & Sederberg, 2014) that were implemented via custom programs in RunDEMC (<https://github.com/compmem/RunDEMC>). Broadly, fitting a model to a set of trial-level data from an individual subject was a 6-step process: First, we specified the relevant model as a system of equations, prior distributions that were determined through a series of pilot investigations, and starting values for each free

parameter. Next, the model was simulated 30,000 times using the starting set of parameter values. This step generated distributions of data in each task condition. The probability density function of the simulated data was then approximated using an Epanechnikov kernel (Turner et al., 2016; Turner & Sederberg, 2014). The estimated functional form of the simulated data approximates the likelihood function of each observed response under the current set of parameter values. The posterior density of the parameter set was calculated as a combination of the likelihood function and the priors. A new proposal parameter set was then selected using differential evolution with Markov chain Monte Carlo (DE-MCMC; Ter Braak, 2006b; Turner et al., 2013; Turner & Sederberg, 2012), a genetic algorithm that selects parameter values according to the success of previous proposals. This procedure was implemented for 800 iterations across 90 chains. Each model's goodness of fit was assessed via Bayesian information criterion (BIC), which considers both the log likelihood of the best-fitting parameters (maximum log likelihood; MLL) and model parsimony (number of free parameters). BIC was calculated using *Equation 4.9*

$$BIC = \ln(n)k - 2\ln(MLL)$$

Equation 4.9

where n is the number of data points and k is the number of free parameters in the model of interest (Claeskens & Hjort, 2001; Wit et al., 2012). BIC values for the circular and elliptical spotlight models were compared within-subject, with lower BIC values indicating a better fit.

4.3 Results

4.3.1 Behavioral Results

Performance scores based on speed and accuracy were calculated for each participant and task condition using Equations 4.1-4.3. As shown in the left panel of Figure 4.3, we observed the expected pattern of decreasing performance from the easy to the moderate condition ($t(25) = 9.65, p < 0.0001$) and from the moderate to the hard condition ($t(25) = 8.74, p < 0.0001$). Performance was also better across participants in the horizontal compared to the vertical condition ($t(25) = 3.92, p < 0.001$). The right panel of Figure 4.3 presents nuanced insight into the latter comparison. The majority (19 out of 26) of participants performed better in the horizontal compared to the vertical condition, but some participants (5 out of 26) displayed the opposite pattern of results. This indicates that most participants were better at ignoring incongruent distractors that were placed immediately above and below the target, compared to those that were placed immediately to the left and right.

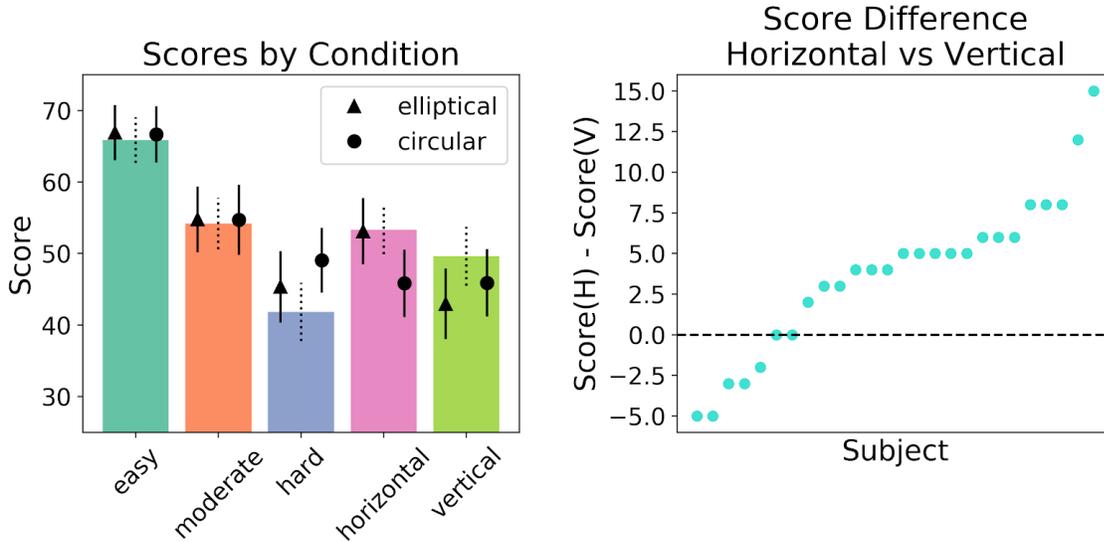


Figure 4.3: Behavioral results. Left Panel: Observed and model-predicted scores within condition. Bars show observed mean scores. Triangular and circular points show mean scores predicted by the elliptical and circular spotlight models, respectively. Error bars show 95% confidence intervals of scores across subjects. Right Panel: Score difference, horizontal vs. vertical conditions. Differences in task scores in the horizontal and vertical task conditions, calculated within-subject.

4.3.2 Model Results

Before assessing our results, we first compared maximum log likelihood (MLL) values from the circular and elliptical spotlight models at the level of each subject as a check for our fitting procedures. Because the elliptical model subsumes the circular variant, the MLL for the elliptical model should be greater than or approximately equal to that of the circular model in all cases. We indeed found this to be true for all subjects. We then calculated differences in BIC values between the two models in order to account for model complexity in addition to MLL when assessing model performance. Results are shown in Figure 4.4. BIC values for the circular model

exceed those of the elliptical model in cases where the addition of the free parameter Θ did not meaningfully improve the fit of the model (6 out of 26 subjects). For 20 out of 26 subjects, however, the additional flexibility of the elliptical model provided meaningfully improved model fits as determined by BIC.

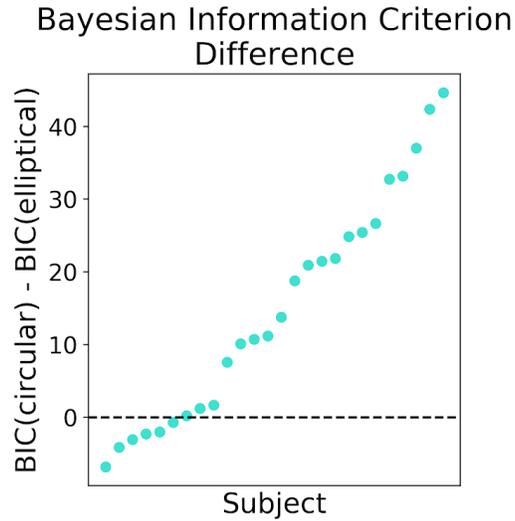


Figure 4.4: Subject-level differences in BIC values between the circular spotlight and elliptical spotlight models. The elliptical model outperforms the circular model for the majority of subjects (lower BIC values indicate better fit).

We next wanted to observe the range of spotlight dimensions calculated within the elliptical model. We first determined the scaled difference between horizontal and vertical standard deviations ($D_{h,v}$) from each subject's best-fitting parameter values in the elliptical spotlight model using Equation 4.10:

$$D_{h,v} = \frac{sd_0(h) - sd_0(v)}{\frac{sd_0(h) + sd_0(v)}{2}}$$

Equation 4.10

where $sd_0(v) = sd_0(h)\Theta$. The left panel of Figure 4.5 shows the range of best-fitting spotlight shapes for each subject: data from 17 out of 26 subjects were best-fit by a horizontally-biased spotlight (points above $y=0.0$), and data from 9 out of 26 subjects were best-fit by a vertically-biased spotlight (points below $y=0.0$). As shown by the right panel of Figure 4.5, we identified a positive correlation between the extent of horizontal bias in spotlight shape and the extent of performance benefit in the horizontal relative to the vertical task condition ($R^2 = 0.31$). The direction of this relationship indicates that participants with a horizontally-biased attentional spotlight are better equipped to ignore distractors placed above and below the target compared to those placed to the left and right of the target, and participants with a vertically-biased spotlight show the opposite performance benefit.

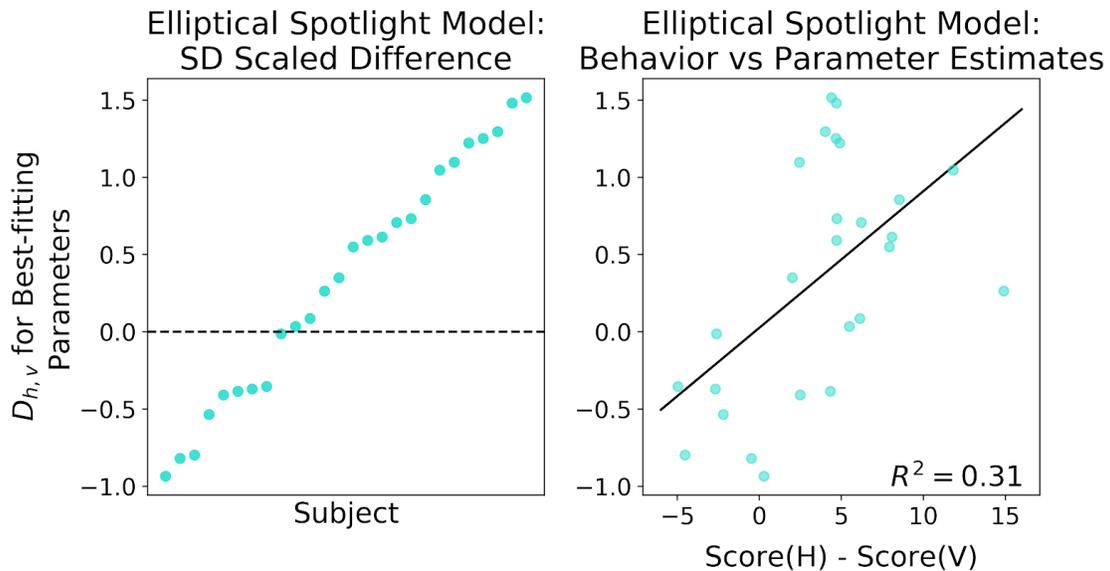


Figure 4.5: Model Results. Left Panel: Subject-level scaled differences in best-fitting horizontal and vertical shape parameters for the spotlight in the elliptical spotlight model. Right Panel: Behavior vs. parameter estimates. Spotlight shape asymmetry predicts behavioral performance in the horizontal and vertical conditions.

To investigate why the elliptical model tended to fit better than the circular variant, we simulated data in each model, and compared them to the behavior that our participants actually produced. We first found best-fitting parameters for each model and subject by identifying the particle in the joint posterior with the maximum weight. Using best-fitting parameters and the relevant model, we generated 10,000 choice-RT trials in each condition. We were then able to calculate a performance score for each model and subject using Equations 4.1-4.3. Ranges of condition-level scores generated by each model are shown in the left panel of Figure 4.3, alongside the observed data. Figure 4.3 shows that both models predict the observed pattern of decreasing performance as we move from the easy, to the moderate, to the hard conditions. The elliptical model also captures the observed pattern of better performance in the moderate and horizontal conditions, compared to the hard and vertical conditions. Given that data from most participants favored a horizontally-biased spotlight (see Figure 4.5), elliptical model predictions reflect the fact that moderate and horizontal stimuli contained the identical configuration of distractors along the horizontal midline, as did the hard and vertical conditions (see Figure 4.1). While the elliptical model made accurate predictions in the moderate and horizontal conditions, it was unable to capture the pattern of better performance in the vertical condition compared to the hard condition. We suspect this is due to perceptual continuity or grouping effects that are not currently accounted for in the model, but that would disproportionately affect performance in the hard condition due to the deliberately “interrupted” configuration of the distractors in the array (Livne & Sagi, 2010; Manassi et al., 2012). Problems arising from a lack of perceptual continuity-related mechanisms appear to be exacerbated by over-constraint of the spotlight dimensions in the circular model. In addition to overestimating performance in the hard

condition, the circular model underestimates performance in both the horizontal and vertical conditions. Because the spotlight in the circular model is constrained to be round, it is unable to predict the observed differences in performance between the horizontal and vertical conditions that are shown in the right panel of Figure 4.3.

4.4 Discussion

Here, we used model-based analyses to investigate individual differences in the shape of the attended visual area while subjects responded to two-dimensional flanker task stimuli. Given evidence of an attentional spotlight that dynamically adjusts to the spatial configuration of stimuli from trial to trial (Pan & Eriksen, 1993), we developed stimulus arrays that were perceptually identical across conditions in order to identify subject-level biases in spotlight dimensions. We constructed two variants of a model that was designed after the zoom lens hypothesis of visual attention (C. Eriksen & St James, 1986; Weichart, Turner, et al., 2020). Consistent with previous results (Andersen & Kramer, 1993; Feng et al., 2007; Hüttermann et al., 2013), we found that most participants use an elliptical spotlight, specifically one with a horizontal bias. This was not the case for all subjects, however, and we identified a widely variable range of spotlight configurations.

Previous studies investigating dimensional biases in visual attention have inferred the shape of the spotlight based on group-level differences in responses to horizontally- and vertically-configured stimuli. To explore individual differences in attentional biases, our model-based methods allowed us to account for the spotlight configuration's critical effects on the cascade of mechanisms that ultimately results in a specific pattern of behavior. Our results therefore add to

the existing literature on attentional allocation, highlighting individual differences in dimensional biases that should be considered in future work. For example, the models in the current study relegated behavioral differences between the horizontal and vertical conditions to differences in spotlight dimensions captured by values of $sd_0(h)$ and $sd_0(v)$. Follow-up work will investigate whether splitting the r_d parameter into separate horizontal and vertical values will further improve model fits, which would indicate that focusing attentional resources occurs at different rates between the two dimensions. Regardless of the precise mechanisms involved, we have presented conclusive evidence of dimension-dependent differences in attentional processes that have not been considered in existing mechanistic models of the flanker task. All other within-trial flanker decision models have been designed to fit data from stimulus arrays oriented along a single horizontal plane (Hübner et al., 2010; Ulrich et al., 2015; White et al., 2011). Real-world challenges to visual attention, however, require integrated processing across multiple spatial dimensions. By failing to consider multidimensional stimuli, our results indicate that these models are potentially missing an important source of variability between individuals in visual processing mechanisms.

Chapter 5: Quantifying Mechanisms of Cognition with an Experiment and Modeling Ecosystem

A major motivating factor in conducting the work presented in Chapters 2-4 was the hope that I would eventually build something that would be helpful for answering clinical questions.

Because reduced performance in cognitive control-related tasks arise alongside several clinical disorders, the ability to pinpoint specific mechanistic deficits through model-based analyses could provide important insight to clinicians when providing diagnoses and treatments. In Chapter 5, I present the model from Chapter 3 and the task from Chapter 4 as tools for fast, detailed assessment of attention and cognitive control. My flanker task and model are part of a 4-piece cognitive task battery, SUPREME, which was developed with my colleagues in Per Sederberg's lab. Chapter 5 contains the details and results of a 2-session validation study, along with examples of using our model-based approach for quantifying cognitive changes through time. The manuscript, presented here in an abridged form to focus on the flanker task, has been submitted to Behavior Research Methods.

5.1 Introduction

Advancements in computational modeling have inspired new ways of studying psychological phenomena, making it possible to quantify elements of the processes underlying behaviors.

Alongside task paradigms designed to target specific cognitive functions, generative models have been used in cognitive psychology for decades to explore mechanistic theories of the computations involved in responding to stimuli. Broadly, generative models fit within a Bayesian framework consist of 1) a system of equations governing the biological processes relevant to the task; 2) a set of free parameters, whose values are responsible for producing different patterns of data by modulating the behavior of the equations; 3) a likelihood function, representing the probability of the data given a set of parameter values; and 4) a set of prior distributions, which specifies the range of plausible values for each parameter (Frässle et al., 2018; Friston et al., 2014; Stephan & Mathys, 2014). The power of generative modeling is that it allows us to mathematically articulate hypotheses about how different layers of processing function and interact, and in turn, formally test said hypotheses by fitting the models to data (Huys et al., 2016). Interpreting behavior in terms of latent processes opens a world of possibilities for identifying the networks involved in complex cognitive functions (Herz et al., 2016; Nunez et al., 2017).

Since the National Institute of Mental Health developed the Research Domain Criteria (RDoC; Insel et al., 2010), generative models have been increasingly prevalent in psychiatric research as well. RDoC is a research framework predicated on the idea that observable symptoms and cognitive behaviors should be studied at all levels of processing, from neural circuits consisting of multiple brain areas down to individual neurons. To this end, several studies have applied model-based analyses to behavior from clinical groups, identifying mechanistic differences in how patients perform on cognitive tasks compared to healthy controls (Moustafa et al., 2015; Pe et al., 2013; C. White et al., 2010, 2016). Several reviews have praised generative modeling as a

potentially transformative tool for psychiatry, but have also noted the problem of balancing specificity (i.e. capturing relevant deficits in individual patient groups) with generalizability (i.e. being applicable in investigations of other patient groups) when developing tasks and models (Adams et al., 2016; Huys et al., 2016; Petzschner et al., 2017). Although a handful of studies have used model-based approaches to target specific dysfunctions relevant to a particular patient group of interest (Cavanagh et al., 2011; Cockburn & Holroyd, 2010; Frank et al., 2007; Mulder et al., 2010; Wiecki et al., 2015) the field currently lacks a comprehensive suite of tasks and associated generative models to measure a full range of cognitive mechanisms across participant groups.

We have developed a toolbox of computerized tasks designed to measure an array of cognitive functions, ranging from low-level perceptual decisions to higher-level assessments of risk, and from working memory to long-term associative memory. Each task in our toolbox is accompanied by a theory-based computational model designed to quantify the latent processes underlying decisions at the level of each trial. We call our toolbox SUPREME: Sensing to Understanding and Prediction Realized via an Experiment and Modeling Ecosystem. SUPREME builds upon existing tasks and models in cognitive and mathematical psychology, respectively. Our goals were to develop a standard, comprehensive means of assessing cognitive functions, and to make mechanistic interpretability of cognitive processes via model-based analyses more accessible for researchers across areas of expertise. Although follow-up work will need to investigate model reliability and applicability to specific diagnostic categories, SUPREME represents a promising step toward a standard method of quantifying and comparing latent cognitive processes across participant groups.

5.1.1 Task Selection

We aimed to create an ecosystem for comprehensive cognitive assessment, provided by a combination of RDoC construct-targeted tasks and mechanistic computational models. In developing SUPREME, we took care to select tasks that 1) are objective, quantifiable measures of the constructs of interest, 2) can be administered multiple times to the same participants with minimal response biases, 3) are simple enough for participants across a wide range of ages and cognitive abilities to complete, 4) are brief, such that each block takes less than five minutes to administer, 5) provide data that are amenable to computational model development and fitting (via constraining task conditions or continuous measures such as RTs in addition to trial-level responses), and 6) span multiple cognitive constructs and different levels of processing complexity. We ultimately selected four tasks that have been mainstays of the cognitive assessment literature for decades, and have been validated by behavioral, neuroimaging, and clinical data to capture group- and condition-level variability in cognitive performance.

Cognitive constructs of interest were selected from RDoC, with a primary focus on the *cognitive systems* domain. The flanker task (Eriksen & Eriksen, 1974; Kopp et al., 1996) was included to measure *attention* and *cognitive control* by requiring participants to make decisions while ignoring distracting information. The task assesses attention by requiring the purposeful focusing of limited visual processing resources, and assesses cognitive control by requiring the suppression of prepotent response modes. In developing our variant of the flanker task, we wanted to create stimuli that were challenging enough to produce differences in performance among cognitively normal participants while still being simple enough for impaired participants

to complete as well. Our variant differs from the classic paradigm in a number of ways. First, target arrows were presented in the center of a 3 by 3-arrow diamond-shaped array, requiring participants to inhibit information in both the horizontal and vertical directions before responding to the target. Second, the configuration of distractor stimuli was based on research demonstrating the *zoom lens* conceptualization of visual attention (Brefczynski & DeYoe, 1999; Tootell et al., 1998). To achieve a gradient of difficulty across task conditions, incongruent distractors could be positioned in either the inner or outer layer of the stimulus array. Third, stimuli were presented on the screen at one of eight possible spatial locations. Participants were therefore required to dynamically modulate their attention on each trial rather than focusing on a single spatial location. The stimuli used in our task variant are illustrated in Figure 5.1. Although I will focus on the flanker task in the context of my dissertation, the other tasks that were included in SUPREME will be briefly described as well.

We selected the random dot motion task (RDM) to measure the construct of *perception*, which encompasses the computations involved in translating sensory input into decision-guiding information. In the task, participants must interpret sensory information in the form of randomly-moving dots and identify the direction of most coherent motion. *Declarative memory*--the encoding, storage, and retrieval of events--was measured with the continuous associative binding task (CAB). The CAB task requires participants to remember relationships between paired items, tapping into associative aspects of declarative memory rather than item recognition alone. Finally, the balloon analogue risk task (BART; Lejuez et al., 2002) was used to assess the *positive* and *negative valence* domains, via the constructs of *reward valuation* and *frustrative nonreward*, respectively. In the task, participants must balance the goal of increasing reward with

the risk of loss by interacting with a virtual balloon. Each task was developed alongside a corresponding computational model, which allowed us to gain insight into the mechanistic underpinnings of the cognitive constructs of interest.

5.1.2 Model Development

An early review by Maia (2015) defined two broad types of model-based analyses in computational psychiatry: 1) *data-driven approaches* in which machine learning methods are used to distinguish among known diagnostic categories of participants, and 2) *theory-driven approaches*, which specify mathematical relationships among variables that contribute to differences in group and individual behaviors. Our methods fall into the latter category, such that we developed sets of algorithms to mathematically describe the processes in between stimulus onset and response. Free parameters in our models represent cognitively interpretable variables, which, when fit to data, provide quantified measures of latent mechanisms that we could not observe from behavior alone. All of our models operate at the level of the individual trial, making it possible to calculate parameter estimates independently for each participant. Models for each task were developed from a combination of existing decision frameworks and mathematical articulations of current cognitive theories. Given that multiple theories exist for how the brain engages certain cognitive processes, our model development procedure was rooted in systematic implementation, fitting, and comparison of contrasting model variants (Kirkpatrick et al., *under review*; Weichart et al., 2020). Experiments were designed and models were developed symbiotically in pursuit of the following goals: 1) performance during experiments should involve the cognitive constructs of interest, 2) models should instantiate the cognitive constructs of interest, and 3) experimental conditions should allow for sufficient constraint, such

that the parameter estimates corresponding to the constructs of interest are informative. Here, ‘informative’ refers to parameter estimates that are both accurate and precise, allowing us to meaningfully differentiate whether a mechanism is a plausible component of an individual’s decision making process. The models presented here provide neurally plausible accounts for the processes underlying the decisions in each task, and also provide the best fits to data compared to alternative accounts as determined by Bayesian comparison analyses.

5.2 Methods

5.2.1 Task

85 participants were recruited to complete the task battery. 1-2 weeks after completing the first session, a subset of 65 participants returned to the lab and completed a second session. Blocks of the four tasks included in SUPREME were randomized within-session.

Participants completed 4 blocks of our flanker task variant, each consisting of 48 trials.

Participants were instructed to indicate the direction of the arrow in the center of the array while ignoring all distractors. Distractor arrows took on one of three configurations (easy, moderate, and hard). On each trial, the stimulus was presented in one of eight locations around the screen.

Possible locations were equidistant from the center of the screen in increments of 45 degrees.

Task condition (easy, moderate, hard), target direction (left or right) and screen location (0, 45, 90, 135, 180, 225, 270, or 315 degrees) were counterbalanced and randomized within-block. The LCA control model from Chapter 3 was fit to data from each subject. Stimulus input to the model was limited to the center horizontal row of each array.

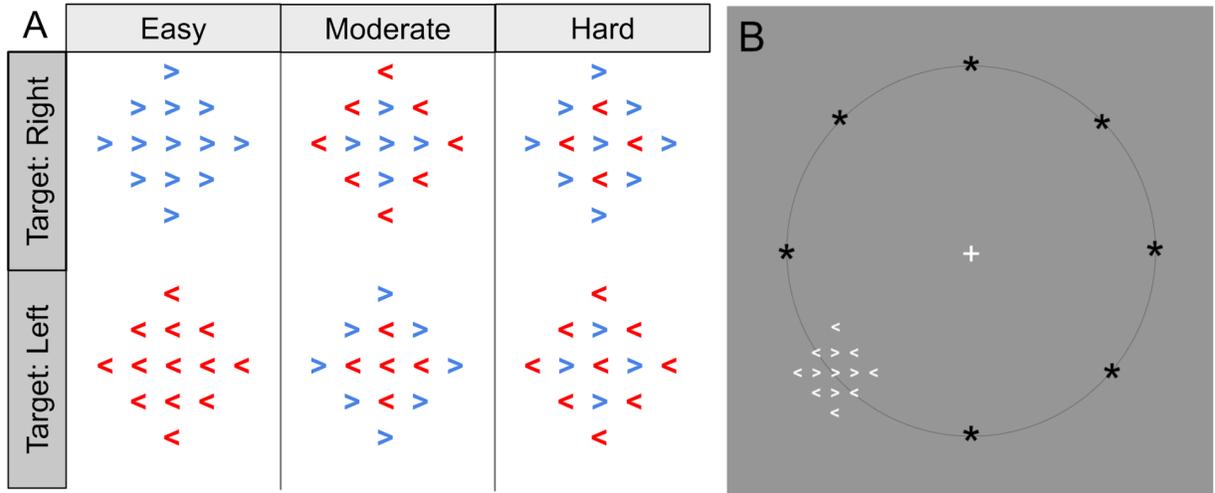


Figure 5.1: Illustration of stimuli in each condition of the flanker task (i.e. easy, moderate, and hard). A) Possible stimulus configurations. Colors were used to highlight the contrasting orientations of the arrows, but stimuli in the actual task were presented in white font against a dark gray background. B) Possible stimulus locations. Black asterisks show the 8 locations in which a stimulus could have appeared on each trial.

5.2.2 Model

Details of the LCA control model as presented in Chapter 3 are summarized here. Because the model is a variant of LCA, it contains free parameters representing decision threshold (α), nondecision time (τ), lateral inhibition (β), and passive decay of evidence (κ). To calculate the drift rate for each choice, we implemented a variation of time-dependent calculations originally described in the SSP (C. White et al., 2011). The SSP draws upon research that points to visual attention behaving as a *zoom lens*, such that perceptual resources are allocated around a central target within a finite area that can expand and contract as needed (Brefczynski & DeYoe, 1999; Mesulam, 1990, 1999; Müller et al., 2003; Tootell et al., 1998). In the model, an attentional spotlight takes the form of a density function for a Gaussian distribution with standard deviation

sd_0 centered upon the central target of the flanker task stimulus. Each arrow in the stimulus array occupies one unit of space and has a perceptual input strength of p . Although our task paradigm features distractor items in both the horizontal and vertical directions, we fit our model based only on the arrows along the horizontal midline of the stimuli for the purposes of our current investigation. In our variant of the model, the standard deviation of the spotlight shrinks as a function of an endogenous calculation of cognitive control, modified by rate of focus (r_d). Our calculation of cognitive control was based on theoretical descriptions of *reactive control* (Braver, 2012; Braver et al., 2008; De Pisapia & Braver, 2006), which suggest that attention is modulated within-trial according to an online calculation of control resources relative to the perceptual conflict within the stimulus. As incongruent stimuli contain evidence for both possible choice options, a greater involvement of attentional mechanisms is required to make a response compared to the case of congruent stimuli. We operationalized the concept of reactive control as a cumulative distance between total evidence and a conflict threshold (δ). Despite neural and behavioral evidence of attention- and cognitive control-mediated changes in the decision process within-trial (Czernochowski, 2015; Nigbur et al., 2015; Scherbaum et al., 2011), SSMS typically calculate evidence as a direct function of time. In a recent model comparison study, we showed that our control-based spotlight implementation within the LCA framework provided better fits to data compared to time-based alternatives, and uniquely mapped onto decision-related signals measured by EEG (Weichart et al., 2020). A list of free parameters and the mechanisms they represent are provided in Table 5.1.

Table 5.1: Summary of flanker model free parameters in Chapter 5

| Parameter | Description |
|-----------|---------------------------|
| α | decision threshold |
| t_0 | nondecision time |
| β | lateral inhibition |
| κ | passive decay of evidence |
| sd_0 | initial spotlight width |
| p | perceptual input strength |
| r_d | rate of focus |
| δ | conflict threshold |

5.3 Results

5.3.1 Behavioral validation

The three conditions of the SUPREME flanker variant indeed provided a more stable gradient of difficulty compared to a standard flanker task, as shown in Figure 5.2. To make a fair comparison between tasks, we selected the first 64 trials per condition from each participant's data in Experiment 1 from Chapter 3 to match the number of trials in the current experiment, and randomly selected first-session data sets from 26 participants in the current experiment to match the number of participants in Experiment 1 from Chapter 3. Figure 5.2A shows that the 95% confidence intervals of condition-level accuracy and speed in the standard flanker task overlapped substantially, with virtually no difference between the congruent and neutral

conditions. In contrast, data from the SUPREME flanker variant in Figure 5.2B shows an expected pattern of decreasing accuracy and increasing mean RTs alongside increasing condition difficulty. We also observe non-overlapping 95% confidence intervals for both accuracy and mean RTs when comparing the easy to the hard condition.

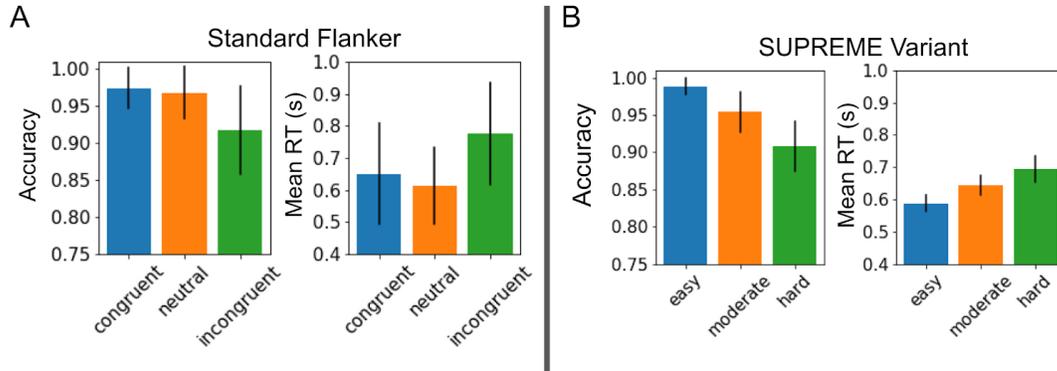


Figure 5.2: Performance differences between a standard flanker paradigm and the SUPREME variant. Bar plots show means and 95% confidence intervals of accuracy and performance across 26 subjects who completed 192 trials of a standard flanker paradigm (A) compared to the flanker task variant included in SUPREME (B).

We calculated standard validation metrics for our SUPREME flanker task variant, using data from all 65 participants who completed both sessions. First, we developed a customized metric on a scale of 0 to 100 that accounts for both speed and accuracy, presented in Chapter 4. In order to earn a high score, participants had to consistently perform quickly and accurately across conditions. Test-retest reliability was “moderate” as determined by the Pearson’s r correlation between session-level metric scores (Devore & Peck, 1993 $r=0.73$), stability of individual differences was “moderate” as determined by Spearman rank order correlation (White, Lejuez, &

de Wit, 2008; $\rho = 0.70$, $p < 0.001$), and internal consistency across four blocks was “high” as determined by Chronbach’s α (Strauss et al., 2006; $\alpha = 0.84$).

5.3.2 Model validation

When we fit the LCA control model presented in Chapter 3 to the data collected using SUPREME, we found that the conditions of the SUPREME flanker variant provided more constraint to parameter estimates compared to those of the standard task. Figure 5.3 shows the posterior distributions for each parameter in our model, mean-centered within-subject and then averaged across 26 subjects from Experiment 1 of Chapter 3 and session 1 of the current study. Despite the fact that the model was fit to data from substantially more trials per subject in the standard flanker task compared to the SUPREME task variant (256 vs. 192 trials), the posterior distributions are more constrained in the latter example. More constrained posteriors mean that we have a greater degree of certainty in our estimates of the parameter values, and the cognitive constructs they represent.

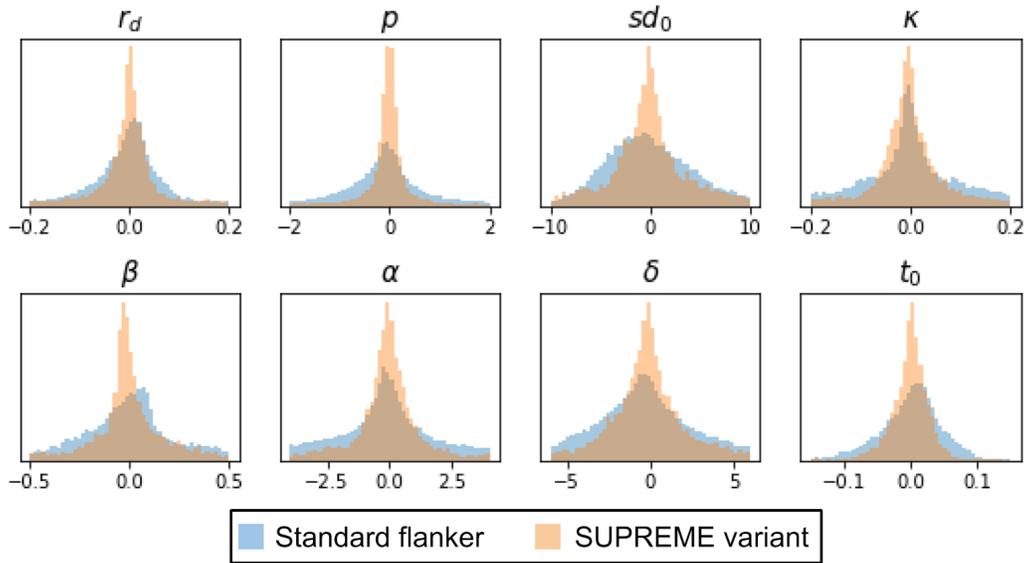


Figure 5.3: Posterior distributions for LCA-control model parameters in Experiments 1 from Chapter 3 and the SUPREME study in Chapter 5. The same model was fit to data from 26 subjects in the standard flanker task and the SUPREME flanker task variant. Parameter values were mean-centered within-subject, and distributions were averaged across subjects.

Qualitative analyses were used to verify that the model was able to capture the observed pattern of behavioral results. We identified a set of best-fitting parameter estimates for each participant's data based on maximum log-likelihood. We generated 10,000 trials within each task condition by inputting each set of parameters back into the model. Figure 5.4A shows means and 95% confidence intervals of scores in each condition, averaged across subjects. For Figure 5.4B, we calculated the Pearson's r correlation between each participant's observed score across conditions and the model-predicted scores generated from each participant's best-fitting parameters ($r=0.82$). From these analyses, we conclude that the model suitably fits the data and generates behavior similarly to our participants.

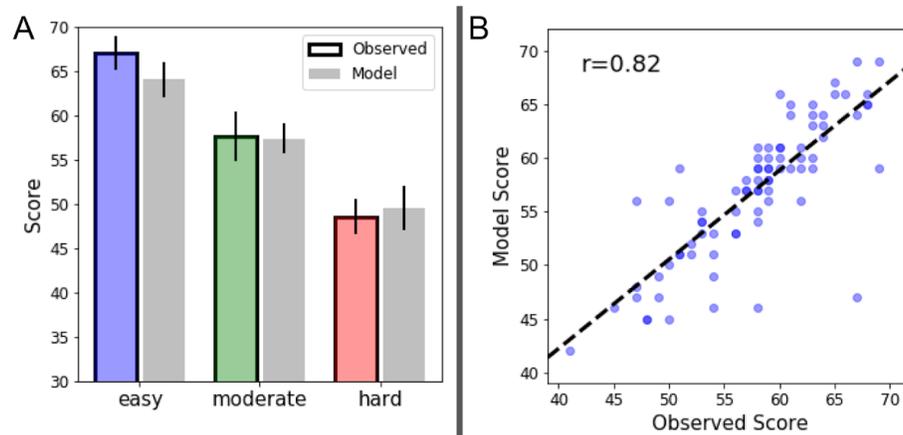


Figure 5.4: SUPREME flanker model results. A) Observed and model-generated scores in each condition of the flanker task. Models were simulated 10,000 times for each condition using each participant's best-fitting parameters. Scores were calculated based on speed and accuracy. Observed mean and 95% CIs of scores are shown as colored bars. Mean and 95% CIs for model-generated scores are shown as gray bars. B) Correlation between observed and model-generated scores across conditions.

5.3.3 Model-based individual difference analyses

The relationship between the brain and behavior is complex. Similar processing modes may lead to different patterns of behavior depending on which brain structures are involved, and the recruitment of different networks could result in nearly identical behavior (Sarter et al., 1996). Compensatory mechanisms are a potent example, whereby the brain upregulates alternative processing routes to maintain cognitive performance despite injury, neuropathology, or otherwise limited resources. At one extreme, several studies of traumatic brain injury patients note cognitive reorganization during rehabilitation, such that entire neural architectures reorient to decrease reliance on dysfunctional connections during cognitive operations (see Galetto &

Sacco, 2017 for review). Among healthy adults, task-related compensatory increases in anterior neural activity have been noted alongside sensory decline resulting from acute cognitive fatigue (Samuel et al., 2019) and as a function of age (Cabeza et al., 2004; Madden et al., 2009). Simple individual differences in processing speed at various levels of neural architecture affect how participants solve a task as well, which may or may not result in differences in overt behavioral performance (Schubert et al., 2019). Analytical methods that are considerate of mechanistic heterogeneity across subjects are therefore important for identifying differences in the processing routes and gaining insight into cognitive states. As a proof of concept, studies have shown that machine learning algorithms trained on best-fitting model parameters rather than data-derived summary statistics are more accurate predictors of young versus old age group membership (Wiecki et al., 2015), as well as diagnostic group classification (Petzschner et al., 2017; Weigard et al., 2020; Wiecki et al., 2016).

One particular strength of our approach is the use of Bayesian analytical procedures, which allow us to identify full distributions of parameter values (the *joint posterior distribution*) that *could* have generated the observed data with some degree of likelihood. As such, we are uniquely positioned to determine both the particular set of values that can optimally recreate a participant's data, as well as a means of quantifying the uncertainty in our measurements of each cognitive construct of interest. Because our models are also *generative* models, meaning they can produce simulated choices and RTs given a set of parameters, we can calculate *posterior predictive distributions* (PPDs) of task scores. PPDs represent the possible ways a participant could have performed, given our model-based assessment of their cognitive state at the time of test. Despite a participant only completing the task set once and producing a single score, PPDs serve as a model-based confidence interval. With the power to quantify the uncertainty in our

measurements via the joint posterior distribution and PPDs, we are able to identify meaningful differences in cognitive acuity at the level of task scores as well as the latent constructs that produced them. As an example, we fit the LCA control model to each participant's first-session data and calculated the joint posterior distribution of parameter values. We then sampled 200 particles from the joint posterior, simulated 10,000 trials in each task condition using each parameter set, and calculated a performance score from each model-generated data set. Figure 5.5 shows the observed scores for two low-performing and two high-performing participants, overlaid with their respective PPDs. Calculations of the percent overlap among the PPDs demonstrate that the two scores from the high-performing participants are meaningfully different from those of the low-performing participants.

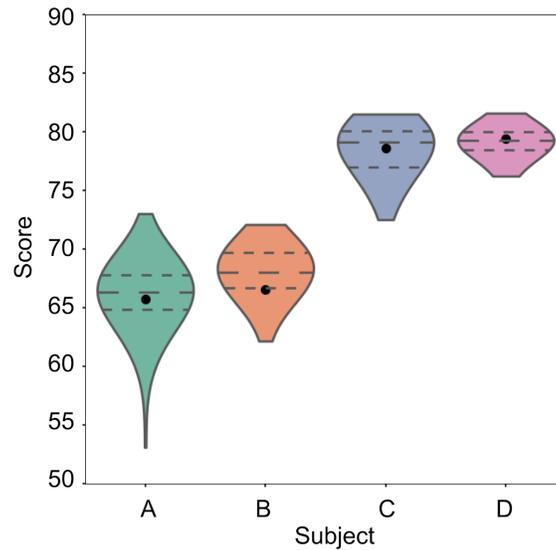


Figure 5.5: Observed scores and PPDs of low-performing (A and B) and high-performing (C and D) participants. Observed scores are displayed as black points, and PPDs are shown as violin plots. Meaningful differences were identified based on percent overlap of PPDs.

Typical behavioral measures such as mean RTs and accuracy provide a sense of each participant's ability to perform the task successfully, but our model-based approach allows us to tap into the nuanced differences in *how* each individual is solving the task. Beyond comparing PPDs, analyses of each participant's joint posterior distribution can reveal nuances in the latent cognitive processes underlying task performance. Figures 5.6 and 5.7 show that although the two pairs of participants from Figure 5.5 ultimately obtained nearly identical scores, their underlying cognitive states were very different. For instance, Figure 5.6 shows that participant B required more evidence before committing to a decision (represented by the decision threshold parameter α) and also committed more attentional resources to processing the flanker stimuli (represented by the initial spotlight width parameter sd_0) compared to participant A. This configuration of parameter values would have perhaps caused participant B to be slower to respond if all other parameters were equal. However, participant B also demonstrated substantially faster visual processing and motor control compared to participant A (represented by the nondecision time parameter t_0), which caused these two participants to ultimately achieve similar scores on the task.

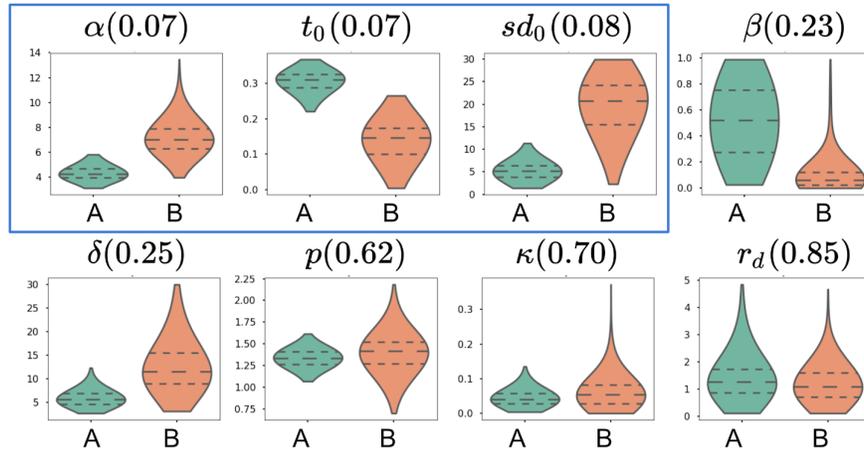


Figure 5.6: Comparison of posterior parameter distributions for two low-performing participants. Facet titles indicate the parameter and the percent overlap between the relevant posterior distributions for participants A and B.

Our assessment of high-performing participants C and D tell a different story, as illustrated in Figure 5.7. Participant C was notably inefficient at focusing the attentional spotlight and filtering out the flanker stimuli on each individual trial, as shown by the rate of focus parameter posterior (r_d). Participant C, however, appears to have compensated for a slow rate of focus via faster visual processing and motor control (t_0) compared to participant D, as well as higher engagement with the task as determined by the general attention parameter (p).

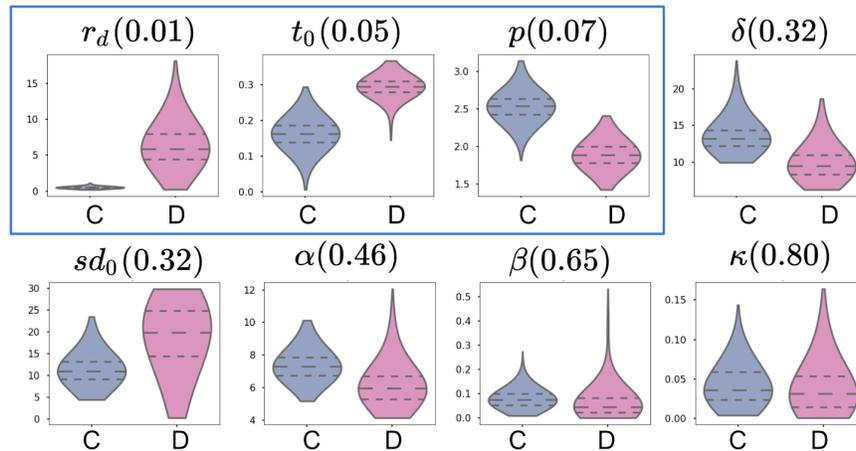


Figure 5.7: Comparison of posterior parameter distributions for two high-performing participants. Facet titles indicate the parameter and the percent overlap between the relevant posterior distributions for participants C and D.

5.4 Discussion

We have developed a battery of cognitive tasks that are simple, objective, quick to administer, and importantly, provide data that are amenable to model development and fitting. Most existing cognitive batteries, which are designed to provide individual metrics that summarize performance, do not provide sufficiently rich data for fitting computational models. In contrast, we developed our tasks and computational models in tandem, defining task conditions that constrain parameter estimates, while capturing fine-grained, individual variation in cognitive abilities. Our tasks -- in combination with our set of models -- provide a means for quickly and objectively evaluating cognitive mechanisms, beyond what we can learn from behavior alone. Here, we have provided validation analyses and model fits for a cohort of cognitively normal participants. In future work, we will extensively investigate how SUPREME can help answer targeted questions, such as how specific parameters relate to variability in neural signals

measured by EEG and fMRI, which parameters are affected by various neuropathologies, and further, how the parameters fluctuate through time according to symptom severity.

Chapter 6: Conclusions

There is a disconnect between what we understand about the biological operations in the brain at the level of the neuron, and the behaviors that serve as the basis of psychiatric diagnoses. As noted by Montague and colleagues (2012), methods for assessing the cognitive functions that operate in between brain and behavior are essential for improving upon individualized treatment plans for psychiatric disorders. My dissertation work serves as an incremental step toward understanding behaviors requiring cognitive control in terms of neural mechanisms, using data from healthy subjects.

The work presented in Chapter 2 investigated the sensitivity of the flanker task for measuring cognitive changes within an individual patient. We administered the flanker task alongside device titration to a patient receiving DBS of the NAcc, a structure known to be involved in functions related to cognitive control (Floresco, 2015). Converging results from behavioral and EEG analyses indicate that acute, significant improvement in task performance during titration serves as an early indicator of subsequent long term treatment effects (Weichart, Sederberg, et al., 2020). In Chapter 3, I investigated competing hypotheses for time-varying decision processes during the flanker task by developing a set of models that varied along dimensions representing drive to attention processes and competition between accumulators. Through three experiments, I found that the model that best-accounted for behavioral and EEG data includes an attentional

spotlight governed by an online calculation of cognitive control, and evidence accumulation mechanisms with leak and lateral inhibition (Weichart, Turner, et al., 2020). In Chapter 4, I utilized the best-performing model from Chapter 3 in an investigation of individual differences in the shape of the attentional spotlight during decisions involving multidimensional stimuli. In Chapter 5, I validated a modified flanker task along with an associated model, and presented methods for between-session assessments using PPDs and joint posterior distributions.

As a whole, my dissertation work seeks to characterize how the healthy brain makes decisions requiring cognitive control. As dysfunctional decision making is central to a wide array of psychiatric disorders, a blueprint of the mechanisms underlying task-related behavior has the potential to be a useful tool for cognitive assessments in the clinical setting. For example, autistic (N. Adams & Jarrold, 2012) and depressed individuals (Ruchow et al., 2004) have both been shown to perform poorly on the flanker task compared to healthy controls. Given the disparate sets of symptoms between the two groups, there may be different cognitive profiles at work that we would be able to quantify with a model-based approach, despite similar overt behavior. In future work throughout my career, I hope to utilize theory-based mechanistic models to investigate clinical questions, understand psychiatric disorders at the level of their component neural processes, and develop new tools for cognitive assessment.

References

- Abbot, L. (1991). Firing rate models for neural populations. In O. Benhar, C. Bosio, P. Giudice, & E. Tabet (Eds.), *Neural networks: From biology to high energy physics* (pp. 179–196). Pisa, Italy: ETS Editrice.
- Adams, N., & Jarrold, C. (2012). Inhibition in autism: children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *Journal of Autism and Developmental Disorders*, *42*(6), 1052–1063.
- Adams, R., Huys, Q., & Roiser, J. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry*, *87*(1), 53–63.
- Amit, D., Brunel, N., & Tsodyks, M. (1994). Correlations of cortical Hebbian reverberations: theory versus experiment. *Journal of Neuroscience*, *14*(11 Pt 1), 6435–6445.
- Andersen, G., & Kramer, A. (1993). Limits of focused attention in three-dimensional space. *Perception & Psychophysics*, *53*(6), 658–667.
- Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, *94*(2), 443–458.
- Awh, E., Anllo-Vento, L., & Hillyard, S. (2000). The role of spatial selective attention in working memory for locations: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*(5), 840–847.

- Baumeister, S., Hohmann, S., Wolf, I., Plichta, M., Rechtsteiner, S., Zangl, M., Ruf, M., Holz, N., Boecker, R., Meyer-Lindenberg, A., Holtmann, M., Laucht, M., Banaschewski, T., & Brandeis, D. (2014). Sequential inhibitory control processes assessed through simultaneous EEG-fMRI. *NeuroImage, 94*, 349–359.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, 57(1)*, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blais, C., Robidoux, S., Risko, E., & Besner, D. (2007). Item-specific adaptation and the conflict-monitoring hypothesis: a computational model. *Psychological Review, 114(4)*, 1076–1086.
- Blond, S., Caparros-Lefebvre, D., Parker, F., Assaker, R., Petit, H., Guieu, J. D., & Christiaens, J. L. (1992). Control of tremor and involuntary movement disorders by chronic stereotactic stimulation of the ventral intermediate thalamic nucleus. *Journal of Neurosurgery, 77(1)*, 62–68.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113(4)*, 700–765.
- Botvinick, M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience, 7(4)*, 356–366.
- Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108(3)*, 624–652. <https://doi.org/10.1037//0033-295x.108.3.624>
- Botvinick, M., Cohen, J., & Carter, C. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences, 8(12)*, 539–546.

- Botvinick, M., Nystrom, L., Fissell, K., Carter, C., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179–181.
- Braver, T. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113.
- Braver, T., Gray, J., & Burgess, G. (2008). Explaining the Many Varieties of Working Memory Variation: Dual Mechanisms of Cognitive Control. In Conway, A., Jarrold, C., Kane, E., Miyake, A., & Towse, J. (Eds.), *Variation in Working Memory* (pp. 76–106). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195168648.003.0004>
- Brefczynski, J., & DeYoe, E. (1999). A physiological correlate of the “spotlight” of visual attention. *Nature Neuroscience*, *2*(4), 370–374.
- Brouillette, R., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C., Bruce-Keller, A., & Keller, J. (2013). Feasibility, Reliability, and Validity of a Smartphone Based Application for the Assessment of Cognitive Function in the Elderly. *PLoS ONE* *8*(6), e65925.
<https://doi.org/10.1371/journal.pone.0065925>
- Brown, J., Reynolds, J., & Braver, T. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology*, *55*(1), 37–85.
- Brown, S., Ratcliff, R., & Smith, P. (2006). Evaluating methods for approximating stochastic differential equations. *Journal of Mathematical Psychology*, *50*(4), 402–410.
- Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336.

- Busch, N., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16048–16053.
- Casaletto, K., & Heaton, R. (2017). Neuropsychological Assessment: Past and Future. *Journal of the International Neuropsychological Society*, *23*(9-10), 778–790.
- Castellanos, N., & Makarov, V. (2006). Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis. *Journal of Neuroscience Methods*, *158*(2), 300–312.
- Castiello, U., & Umiltà, F. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 837–848.
- Chen, C.-C., & Tyler, C. (2002). Lateral modulation of contrast discrimination: flanker orientation effects. *Journal of Vision*, *2*(6), 520–530.
- Chen, G., Saad, Z., Britton, J., Pine, D., & Cox, R. (2013). Linear mixed-effects modeling approach to fMRI group analysis. *NeuroImage*, *73*, 176–190.
- Churchland, A., Kiani, R., & Shadlen, M. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702.
- Claeskens, G., & Hjort, N.-L. (2001). *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics: Cambridge University Press.
- Cohen, A., & Shoup, R. (1993). Orientation asymmetry in the flanker task. *Perception & Psychophysics*, *53*(6), 693–703.
- Colonus, H., & Diederich, A. (2018). Paradox resolved: Stop signal race model with negative dependence. *Psychological Review*, *125*(6), 1051-1058. <https://doi.org/10.1037/rev0000127>

- Czernochowski, D. (2015). ERPs dissociate proactive and reactive control: evidence from a task-switching paradigm with informative and uninformative cues. *Cognitive, Affective & Behavioral Neuroscience, 15*(1), 117–131.
- Davelaar, E. (2008). A computational study of conflict-monitoring at two levels of processing: reaction time distributional analyses and hemodynamic responses. *Brain Research, 1202*, 109–119.
- De Jong, R., Liang, C., & Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20*(4), 731–750.
- De Pisapia, N., & Braver, T. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing, 69*(10-12), 1322–1326.
<https://doi.org/10.1016/j.neucom.2005.12.100>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.
- Devore, J., & Peck, R. (1993). *Statistics: Exploration and Analysis*. Arden Shakespeare.
- Ditterich, J. (2010). A Comparison between Mechanisms of Multi-Alternative Perceptual Decision Making: Ability to Explain Human Behavior, Predictions for Neurophysiology, and Relationship with Decision Theory. *Frontiers in Neuroscience, 4*, 184.
- Dubois, J., Hamker, F., & VanRullen, R. (2009). Attentional selection of noncontiguous locations: the spotlight is only transiently “split.” *Journal of Vision, 9*(5), 3.1–11.
- Eriksen, B., & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149.

- Eriksen, C., & Schultz, D. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Perception & Psychophysics*, *25*(4), 249–263.
- Eriksen, C., & St James, J. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.
- Eriksen, C., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology. Human Perception and Performance*, *11*(5), 583–597.
- Ezzyat, Y., Kragel, J., Burke, J., Levy, D., Lyalenko, A., Wanda, P., O’Sullivan, L., Hurley, K., Busygin, S., Pedisich, I., Sperling, M., Worrell, G., Kucewicz, M., Davis, K., Lucas, T., Inman, C., Lega, B., Jobst, B., Sheth, S., Zaghloul, K., Jutras, M., Stein, J., Das, S., Gorniak, R., Rizzuto, D., & Kahana, M. J. (2017). Direct brain stimulation modulates encoding states and memory performance in humans. *Current Biology*, *27*(9), 1251–1258.
- Feng, C., Jiang, Y., & He, S. (2007). Horizontal and vertical asymmetry in visual spatial crowding effects. *Journal of Vision*, *7*(2), 13.
- Figee, M., Luigjes, J., Smolders, R., Valencia-Alfonso, C., Van Wingen, G., De Kwaasteniet, B., Mantione, M., Ooms, P., De Koning, P., Vulink, N., & Levar, N. (2013). Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. *Nature Neuroscience*, *16*(4), 386–387.
- Floresco, S. (2015). The nucleus accumbens: An interface between cognition, emotion, and action. *Annual Review of Psychology*, *66*, 25–52.
- Formolo, D., Gaspar, J., Melo, H., Eichwald, T., Zepeda, R., Latini, A., Okun, M., & Walz, R. (2019). Deep Brain Stimulation for Obesity: A Review and Future Directions. *Frontiers in Neuroscience*, *13*, 323.

- Frank, M. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks: The Official Journal of the International Neural Network Society*, *19*(8), 1120–1136.
- Fujii, D. (2017). *Conducting a culturally informed neuropsychological evaluation*. Washington, DC, US: American Psychological Association.
- Goschke, T., & Dreisbach, G. (2008). Conflict-triggered goal shielding: response conflicts attenuate background monitoring for prospective memory cues. *Psychological Science*, *19*(1), 25–32.
- Goto, Y., & Grace, A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience*, *8*(6), 805–812.
- Grace, A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, *95 Suppl 2*, S119–S128.
- Gratton, G., Coles, M., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology. General*, *121*(4), 480–506.
- Gratton, G., Coles, M., Sirevaag, E., Eriksen, C., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(3), 331–344.
- Haider, M., Spong, P., & Lindsley, D. (1964). Attention, Vigilance, and Cortical Evoked-Potentials in Humans. *Science*, *145*(3628), 180–182.
- Handy, T., Soltani, M., & Mangun, G. (2001). Perceptual load and visuocortical processing: event-related potentials reveal sensory-level selection. *Psychological Science*, *12*(3), 213–218.
- Harat, M., Rudaś, M., Zieliński, P., Birska, J., & Sokal, P. (2016). Nucleus accumbens stimulation in pathological obesity. *Neurologia I Neurochirurgia Polska*, *50*(3), 207–210.

- Heaton, R., Grant, I., & Matthews, C. (1991). *Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographic Corrections, Research Findings, and Clinical Applications*.
- Heinze, H.-J., Heldmann, M., Voges, J., Hinrichs, H., Marco-Pallares, J., Hopf, J.-M., Müller, U., Galazky, I., Sturm, V., Bogerts, B., & Münte, T. (2009). Counteracting incentive sensitization in severe alcohol dependence using deep brain stimulation of the nucleus accumbens: Clinical and basic science aspects. *Frontiers in Human Neuroscience*, 3, 22.
- Heinze, H., Mangun, G., Burchert, W., Hinrichs, H., Scholz, M., Münte, T., Gös, A., Scherg, M., Johannes, S., & Hundeshagen, H. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, 372(6506), 543–546.
- Henderson, J. (1991). Stimulus discrimination following covert attentional orienting to an exogenous cue. *Journal of Experimental Psychology. Human Perception and Performance*, 17(1), 91–106.
- Hillyard, S., Teder-Sälejärvi, W., & Münte, T. (1998). Temporal dynamics of early perceptual processing. In *Current Opinion in Neurobiology* (Vol. 8, Issue 2, pp. 202–210).
[https://doi.org/10.1016/s0959-4388\(98\)80141-4](https://doi.org/10.1016/s0959-4388(98)80141-4)
- Holdnack, J., Tulskey, D., Brooks, B., Slotkin, J., Gershon, R., Heinemann, A., & Iverson, G. (2017). Interpreting Patterns of Low Scores on the NIH Toolbox Cognition Battery. *Archives of Clinical Neuropsychology*, 32(5), 574–584.
- Holmes, W. (2015). A practical guide to the Probability Density Approximation (PDA) with improved implementation and error characterization. *Journal of Mathematical Psychology*, 68-69, 13–24.
- Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, 117(3), 759–784.

- Hunt, L., Malalasekera, W., de Berker, A., Miranda, B., Farmer, S., Behrens, T., & Kennerley, S. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience*, *21*(10), 1471–1481.
- Hüttermann, S., Memmert, D., & Simons, D. (2014). The size and shape of the attentional “spotlight” varies with differences in sports expertise. *Journal of Experimental Psychology: Applied*, *20*(2), 147–157.
- Hüttermann, S., Memmert, D., Simons, D., & Bock, O. (2013). Fixation strategy influences the ability to focus attention on two spatially separate objects. *PloS One*, *8*(6), e65673.
- Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience and Biobehavioral Reviews*, *46 Pt 1*, 30–43.
- Jonides, J. (1983). Further toward a model of the Mind’s eye’s movement. *Bulletin of the Psychonomic Society*, *21*(4), 247–250.
- Kalambet, Y., Kozmin, Y., & Samokhin, A. (2018). Comparison of integration rules in the case of very narrow chromatographic peaks. *Chemometrics and Intelligent Laboratory Systems*, *179*, 22–30. <https://doi.org/10.1016/j.chemolab.2018.06.001>
- Kerns, J., Cohen, J., MacDonald, A., Cho, R., Stenger, V., & Carter, C. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023–1026.
- Kirkpatrick, R., Turner, B., & Sederberg, P. (under review). *Equal evidence perceptual tasks suggest key role for interactive competition in decision-making.*
- Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*(3), 282–294.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review*, *97*(2), 253–270.

- LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, 9(3), 371–379.
- Laming, D. (1968). *Information theory of choice-reaction times*. Academic Press Inc.
- Larson, M., Clayson, P., & Clawson, A. (2014). Making sense of all the conflict: a theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology*, 93(3), 283–297.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 451–468.
- Le, D., Pannacciulli, N., Chen, K., Del Parigi, A., Salbe, A., Reiman, E., & Krakoff, J. (2006). Less activation of the left dorsolateral prefrontal cortex in response to a meal: A feature of obesity. *The American Journal of Clinical Nutrition*, 84(4), 725–731.
- Liu, Y., Holmes, P., & Cohen, J. (2008). A neural network model of the Eriksen task: reduction, analysis, and data fitting. *Neural Computation*, 20(2), 345–373.
- Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, 10(3), 1.1–14.
- Livne, T., & Sagi, D. (2011). Multiple levels of orientation anisotropy in crowding with Gabor flankers. *Journal of Vision*, 11(13), 18.
- Luck, S., Hillyard, S., Mangun, G., & Gazzaniga, M. (1989). Independent hemispheric attentional systems mediate visual search in split-brain patients. *Nature*, 342(6249), 543–545.
- MacDonald, A., Cohen, J., Stenger, V., & Carter, C. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288(5472), 1835–1838.

- Mack, M., Preston, A., & Love, B. (2013). Decoding the Brain's Algorithm for Categorization from Its Neural Implementation. *Current Biology*, *23*(20), 2023-2027.
<https://doi.org/10.1016/j.cub.2013.08.035>
- Maia, T., Huys, Q., & Frank, M. (2017). Theory-Based Computational Psychiatry. *Biological Psychiatry*, *82*(6), 382-384.
- Manassi, M., Sayim, B., & Herzog, H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10), 13.
- Mangun, G., & Hillyard, S. (1988). Spatial gradients of visual attention: behavioral and electrophysiological evidence. *Electroencephalography and Clinical Neurophysiology*, *70*(5), 417-428.
- Manly, J., Jacobs, D., Touradji, P., Small, S., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society: JINS*, *8*(3), 341-348.
- Mantione, M., van de Brink, W., Schuurman, P. R., & Denys, D. (2010). Smoking cessation and weight loss after chronic deep brain stimulation of the nucleus accumbens: therapeutic and research implications: case report. *Neurosurgery*, *66*(1), E218; discussion E218.
- Mayr, U., & Awh, E. (2009). The elusive link between conflict and conflict adaptation. *Psychological Research*, *73*(6), 794-802. <https://doi.org/10.1007/s00426-008-0191-1>
- McClelland, J., & Cleeremans, A. (2009). Consciousness and connectionist models. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *The Oxford companion to consciousness*. Oxford University Press.
- McIntyre, C., Savasta, M., Walter, B., & Vitek, J. L. (2004). How does deep brain stimulation work? Present understanding and future questions. *Journal of Clinical Neurophysiology*, *21*(1), 40-50.

- Mesulam, M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5), 597–613.
- Mesulam, M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1387), 1325–1346.
- Mikos, A., Bowers, D., Noecker, A., McIntyre, C., Won, M., Chaturvedi, A., Foote, K., & Okun, M. (2011). Patient-specific analysis of the relationship between the volume of tissue activated during DBS and verbal fluency. *NeuroImage*, 54 Suppl 1, S238–S246.
- Montague, P., Dolan, R., Friston, K., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72–80.
- Müller, M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked potentials. *Psychological Science*, 13(2), 119–124.
- Müller, M., Malinowski, P., Gruber, T., & Hillyard, S. (2003). Sustained division of the attentional spotlight. *Nature*, 424(6946), 309–312.
- Müller, N., Bartelt, O., Donner, T., Villringer, A., & Brandt, S. (2003). A Physiological Correlate of the “Zoom Lens” of Visual Attention. *The Journal of Neuroscience*, 23(9), 3561–3565.
- Müller, N., Mollenhauer, M., Rösler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research*, 45(9), 1129–1137.
- Nigbur, R., Schneider, J., Sommer, W., Dimigen, O., & Stürmer, B. (2015). Ad-hoc and context-dependent adjustments of selective attention in conflict control: an ERP study with visual probes. *NeuroImage*, 107, 76–84.

- Norman, D., & Shallice, T. (1986). Attention to Action. In *Consciousness and self-regulation* (pp. 1–18). Springer.
- Nuttin, B., Cosyns, P., Demeulemeester, H., Gybels, J., & Meyerson, B. (1999). Electrical stimulation in anterior limbs of internal capsules in patients with obsessive-compulsive disorder. *The Lancet*, *354*(9189), 1526.
- Palestro, J., Sederberg, P., Osth, A., Van Zandt, T., & Turner, B. (2018). Likelihood-Free Methods for Cognitive Science. In *Computational Approaches to Cognition and Perception*.
<https://doi.org/10.1007/978-3-319-72425-6>
- Panagopoulos, A., Von Grünau, M., & Galera, C. (2004). Attentive mechanisms in visual search. *Spatial Vision*, *17*(4-5), 353–371.
- Pan, K., & Eriksen, C. (1993). Attentional distribution in the visual field during same-different judgments as assessed by response competition. *Perception & Psychophysics*, *53*(2), 134–144.
- Parsons, T. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience*, *9*, 660.
- Passamonti, L., Rowe, J., Schwarzbauer, C., Ewbank, M., von dem Hagen, E., & Calder, A. (2009). Personality predicts the brain's response to viewing appetizing foods: the neural basis of a risk factor for overeating. *The Journal of Neuroscience*, *29*(1), 43–51.
- Peirce, J. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.
- Philiastides, M., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *The Journal of Neuroscience*, *26*(35), 8965–8975.

- Ram, A., Jalal, S., Jalal, A., & Kumar, M. (2010). A density based algorithm for discovering density varied clusters in large spatial databases. *International Journal of Computer Applications in Technology*, 3(6), 1–4.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59-108.
<https://doi.org/10.1037//0033-295x.85.2.59>
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106(2), 261–300.
- Rezai, A., Krishna, V., Bogner, J., Kramer, D., Needleman, B., Emerson, A., Reider, C., Fiore, G., & Corrigan, J. (2018). Letter: Feasibility of Nucleus Accumbens Deep Brain Stimulation for Morbid, Treatment-Refractory Obesity. *Neurosurgery*, 82(5), E136–E137.
<https://doi.org/10.1093/neuros/nyx630>
- Ridderinkhof, K. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research*, 66(4), 312–323.
- Ridderinkhof, K., Ullsperger, M., Crone, E., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306(5695), 443–447.
- Ruchow, M., Herrnberger, B., Wiesend, C., Grön, G., Spitzer, M., & Kiefer, M. (2004). The effect of erroneous responses on response monitoring in patients with major depressive disorder: a study with event-related potentials. *Psychophysiology*, 41(6), 833–840.
- Scherbaum, S., Dshemuchadse, M., Ruge, H., & Goschke, T. (2012). Dynamic goal states: adjusting cognitive control without conflict monitoring. *NeuroImage*, 63(1), 126–136.
- Scherbaum, S., Fischer, R., Dshemuchadse, M., & Goschke, T. (2011). The dynamics of cognitive control: evidence for within-trial conflict adaptation from frequency-tagged EEG. *Psychophysiology*, 48(5), 591–600.

- Servant, M., Gajdos, T., & Davranche, K. (2018). ELF: A new measure of response capture. *Psychonomic Bulletin & Review*, 25(2), 539–547.
- Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, 72, 162–195.
- Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using Covert Response Activation to Test Latent Assumptions of Formal Decision-Making Models in Humans. *The Journal of Neuroscience*, 35(28), 10371–10385.
- Sesia, T., Bulthuis, V., Tan, S., Lim, L., Vlamings, R., Blokland, A., Steinbusch, H., Sharp, T., Visser-Vandewalle, V., & Temel, Y. (2010). Deep brain stimulation of the nucleus accumbens shell increases impulsive behavior and tissue levels of dopamine and serotonin. *Experimental Neurology*, 225(2), 302–309.
- Shadlen, M., & Newsome, W. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916–1936.
- Smith, P., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168.
- Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98.
- Strauss, E., Sherman, E., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. American Chemical Society.
- Sturm, V., Lenartz, D., Koulousakis, A., Treuer, H., Herholz, K., Klein, J., & Klosterkötter, J. (2003). The nucleus accumbens: A target for deep brain stimulation in obsessive–compulsive- and anxiety-disorders. *Journal of Chemical Neuroanatomy*, 26(4), 293–299.

- Ter Braak, C. (2006a). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3).
- Tootell, R., Hadjikhani, N., Hall, E., Marrett, S., Vanduffel, W., Vaughan, J., & Dale, A. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*(6), 1409–1422.
- Treue, S., & Martinez-Trujillo, J. (2012). The spotlight of attention: shifting, resizing and splitting receptive fields when processing visual motion. *E-Neuroforum*, *18*(3).
<https://doi.org/10.1007/s13295-012-0034-9>
- Tronnier, V., Rasche, D., Thorns, V., Alvarez-Fischer, D., Münte, T. F., & Zurovski, B. (2018). Massive weight loss following deep brain stimulation of the nucleus accumbens in a depressed woman. *Neurocase*, *24*(1), 49–53.
- [Turner, B., Forstmann, B., Love, B., Palmeri, T., & Van Maanen, L. \(2017\). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. <https://doi.org/10.1016/j.jmp.2016.01.001>](https://doi.org/10.1016/j.jmp.2016.01.001)
- Turner, B., Schley, D., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329–362.
- Turner, B., & Sederberg, P. (2012). Approximate Bayesian computation with differential evolution. *Journal of Mathematical Psychology*, *56*(5), 375–385.
- Turner, B., & Sederberg, P. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227–250.
- Turner, B., Sederberg, P., Brown, S., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384.
- Turner, B., Sederberg, P., & McClelland, J. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, *72*, 191–199.

- Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.
- Umiltà, F. (1998). The Control of the Attentional Focus. *European Journal of Cognitive Psychology*, *10*(3), 225–246.
- Usher, M., & McClelland, J. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.
- Usher, M., & McClelland, J. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769.
- van Ravenzwaaij, D., van der Maas, H., & Wagenmakers, E.-J. (2012). Optimal decision making in neural inhibition models [Review of *Optimal decision making in neural inhibition models*]. *Psychological Review*, *119*(1), 201–215.
- VanRullen, R., & Thorpe, S. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.
- van Veen, V., & Carter, C. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*(4-5), 477–482.
- Vejnović, D., & Zdravković, S. (2015). Side flankers produce less crowding, but only for letters. *Cognition*, *143*, 217–227.
- Verguts, T. (2017). Binding by Random Bursts: A Computational Model of Cognitive Control. *Journal of Cognitive Neuroscience*, *29*(6), 1103–1118.
- Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological Review*, *115*(2), 518–525.

- Voges, J., Müller, U., Bogerts, B., Münte, T., & Heinze, H.-J. (2013). Deep brain stimulation surgery for alcohol addiction. *World Neurosurgery*, *80*(3-4), S28.e21–e31.
- Volkman, J., Herzog, J., Kopper, F., & Deuschl, G. (2002). Introduction to the programming of deep brain stimulators. *Movement Disorders: Official Journal of the Movement Disorder Society*, *17 Suppl 3*, S181–S187.
- Volkman, J., Moro, E., & Pahwa, R. (2006). Basic algorithms for the programming of deep brain stimulation in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *21 Suppl 14*, S284–S289.
- Volkow, N., Wang, G.-J., Fowler, J., & Telang, F. (2008). Overlapping neuronal circuits in addiction and obesity: Evidence of systems pathology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1507), 3191–3200.
- Ward, R., & Ward, R. (2006). Cognitive conflict without explicit conflict monitoring in a dynamical agent. In *Neural Networks* (Vol. 19, Issue 9, pp. 1430–1436).
<https://doi.org/10.1016/j.neunet.2006.08.003>
- Weichart, E., Darby, K., Fenton, A., Jacques, B., Kirkpatrick, R., Turner, B., & Sederberg, P. (in prep). *Quantifying mechanisms of cognition with an experiment and modeling ecosystem*.
- Weichart, E., Sederberg, P., Sammartino, F., Krishna, V., Corrigan, J., & Rezai, A. (2020). Cognitive task performance during titration predicts deep brain stimulation treatment efficacy: Evidence from a case study. *Frontiers in Psychiatry*, *11*(30). <https://doi.org/10.3389/fpsy.2020.00030>
- Weichart, E., Turner, B., & Sederberg, P. (2020a). A model of dynamic, within-trial conflict resolution for decision making. *Psychological Review*, *Advance online publication*.
<https://doi.org/10.1037/rev0000191>

- White, C., Ratcliff, R., & Starns, J. (2011). Diffusion models of the flanker task: discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238.
- White, C., Servant, M., & Logan, G. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review*, *25*(1), 286–301.
- Wiecki, T., Poland, J., & Frank, M. (2015). Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry. In *Clinical Psychological Science* (Vol. 3, Issue 3, pp. 378–399). <https://doi.org/10.1177/2167702614565359>
- Wild-Wall, N., Oades, R., Schmidt-Wessels, M., Christiansen, H., & Falkenstein, M. (2009). Neural activity associated with executive functions in adolescents with attention-deficit/hyperactivity disorder (ADHD). *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *74*(1), 19–27.
- Wit, E., van den Heuvel, E., & Romeijn, J.-W. (2012). “All models are wrong...”: an introduction to model uncertainty. *Statistica Neerlandica*, *66*(3), 217–236.
- Yao, Y., Frässle, S., Schöbi, D., Aponte, E., Heinzle, J., & Stephan, K. (2018). Generative models for clinical applications in computational psychiatry. *Cognitive Science*, *9*(3), e1460.
- Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165.
- Yeung, N., Botvinick, M., & Cohen, J. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.

Zhu, J., & Tulskey, D. (2000). Co-norming the WAIS-III and WMS-III: Is there a test-order effect on IQ and memory scores? *The Clinical Neuropsychologist*, *14*(4), 461–467.

Appendix A: Supplementary Materials for Chapter 2

A.1 Additional Results

Mean and standard deviations of RTs within the stimulation parameter sets of interest across titration sessions were as follows:

- 1) Lower-middle contacts, low V (optimal): mean=0.582s, sd=0.182s
- 2) Both middle contacts, high V: mean=0.592s, sd=0.182s
- 3) Lower-middle contacts, high V: mean=0.599s, sd=0.151s
- 4) DBS-OFF: mean=0.633s, sd=0.182s

Results comparing correct RTs in the incongruent condition during stimulation with optimal stimulation parameter values (low V, ventral-medial contacts) to either set of suboptimal parameter values were not statistically significant:

- 1) Lower-middle contacts, low V (optimal) vs. both middle contacts, high V: $X^2(1)=2.276$,
p=0.131
- 2) Lower-middle contacts, low V (optimal) vs. lower-middle contacts, high V: $X^2(1)=1.164$,
p=0.281

Results comparing correct RTs in the congruent condition during stimulation with optimal parameter values to either set of suboptimal parameter values were also not significant:

- 1) Lower-middle contacts, low V (optimal) vs. both middle contacts, high V: $X^2(1)=0.327$,
p=0.567

- 2) Lower-middle contacts, low V (optimal) vs. lower-middle contacts, high V: $X^2(1)=0.167$, $p=0.682$

Results comparing correct RTs in the neutral condition during stimulation for each parameter set to DBS-OFF were not significant:

- 1) Lower-middle contacts, low V (optimal) vs. DBS-OFF: $X^2(1)=1.972$, $p=0.160$
- 2) Both middle contacts, high V vs. DBS-OFF: $X^2(1)=0.166$, $p=0.684$
- 3) Lower-middle contacts, high V vs. DBS-OFF: $X^2(1)=0.359$, $p=0.543$

Long-term results were presented as rate of weight loss, calculated by dividing the total weight change by the total number of days that each set of stimulation parameters were active. As each set of parameters were active for very different amounts of time (129, 675, and 108 days), our rate calculations may have been biased by duration-dependent changes in stimulation efficacy. In a follow-up analysis, we therefore considered weight loss only within the initial period of active stimulation in an effort to standardize these long-term durations. Although durations were not exactly equal because we did not collect weight data on every day of the study, we find that the results presented in the main manuscript are maintained:

- 1) Lower-middle contacts, low V: 47.8 lbs. lost in 129 days
- 2) Both middle contacts, high V: 22.2 lbs. lost in 119 days
- 3) Lower-middle contacts, high V: 1.6 lbs. lost in 108 days

For the main analyses, voltages were assigned to “low” and “high” groups relative to 5V. To demonstrate that our results are maintained beyond these specific group designations, we ran follow-up analyses with three voltage groups, each covering an equal range of 2.67V: low (2.00 to 4.67V), medium (4.67 to 7.33V), and high (7.33 to 10.00V). Using the same linear mixed effects regression analysis described in the manuscript, we identified a significant flanker RT

effect for only one set of stimulation parameters: bilateral lower-middle contacts with medium V in the left hemisphere (L) and low V in the right hemisphere (R; $X^2(1)=6.350$, $p=0.012$).

Applying these group delineations to the long-term weight loss data, we found that the same set of parameters that yielded a significant acute task effect also yielded the fastest rate of weight loss in the long term:

- 1) Lower-middle contacts, L medium V, R low V: 45.8 lbs. lost in 121 days

Another set of parameters yielded the same amount of weight loss, but over a much longer period of time:

- 2) Both middle contacts, L high V, R high V: 45.8 lbs. lost in 611 days

Other results were as follows:

- 3) Lower-middle contacts, L low V, R low V: 2.0 lbs. lost in 8 days
- 4) Lower-middle contacts, L medium V, R medium V: 5.7 lbs. gained in 14 days
- 5) Lower-middle contacts, L high V, R high V: 7.3 lbs. lost in 94 days
- 6) Both middle contacts, L medium V, R medium V: 1.2 lbs. lost in 64 days

A.2 DTI Connectivity

DTI connectivity analyses were conducted by my collaborators, Drs. Francesco Sammartino and Vibhor Krishna, and their results were included in the version of the manuscript that was published in *Frontiers in Psychiatry*. As I did not conduct the connectivity analyses myself, I excluded the associated methods and results from Chapter 2. I do believe, however, that reference to the connectivity results are critical for interpreting the behavioral and EEG findings as presented in Section 2.4.2. Methods and results of the connectivity analyses are therefore provided in the sections to follow.

A.2.1 Methods

Patterns of structural connectivity were assessed using preoperative diffusion weighted imaging (60 diffusion directions, 2 mm iso-voxel, Philips Ingenia CX). The structural (T1-weighted and post-operative computed tomography images) and diffusion weighted images were co-registered and nonlinearly aligned to the MNI152 T1 template space 0.5mm. Despite only analyzing data from one subject, we co-registered images to MNI space in order to overcome the limitations of morphing standard network atlases to a single brain, as well as provide results in an interpretable, universal space. The volumes of tissue activation were calculated for each stimulation parameter set and electrode impedance using methods described by Vorwerk and colleagues (2013; also in Horn et al., 2017). The volumes of tissue activation for the same three sets of stimulation parameters used in our behavioral analyses (optimal: bilateral lower middle contacts, low amplitudes, suboptimal: bilateral lower middle contacts, high amplitudes and both bilateral middle contacts, high amplitudes) were used as seeds for probabilistic tractography (FSL, three fibers per voxel model, 25000 samples per voxel). A validated cortical multimodality atlas was used to generate target masks (Yeo et al., 2011). The connectivity maps resulting from each individual voltage change were thresholded to the 99 percentile of 'robust' intensity and were subjected to a linear mixed-effects model using AFNI's 3dMVM program (Chen, Saad, Britton, Pine, & Cox, 2013) with the amplitude of stimulation as the within-subject variable. The resulting statistical maps thresholded by the false discovery rate corrected p-value of $<.01$ were clusterized using an unsupervised density-based clustering algorithm (DBSCAN-R library; Ram, Jalal, Jalal, & Kumar, 2010) with parameters epsilon 1.6 and 5 minimum points. The measure of connectivity and directionality (i.e increased vs. decreased connectivity) was interpreted based

on the number of voxels intersecting the network masks, following a methodology reported by Klein and colleagues (2012).

A.2.2 Results

Volumes of tissue activation associated with three stimulation settings (optimal: bilateral lower middle contacts, low amplitudes; suboptimal: bilateral lower middle contacts, high amplitudes and both bilateral middle contacts, high amplitudes) were used as seeds for probabilistic tractography with target masks derived from a multimodality cortical atlas. Probability of connectivity was determined based on the number of voxels intersected by tractography in each network mask in the three settings groups. Optimal DBS settings were associated with higher probability of connectivity with right dorsolateral and dorsomedial prefrontal cortex as shown in Figure A1, panels A-D. When comparing the optimal versus suboptimal settings, the significant cortical clusters were localized within frontal regions (basal frontal, cingulate) as shown in Figure A1, panels E-H.

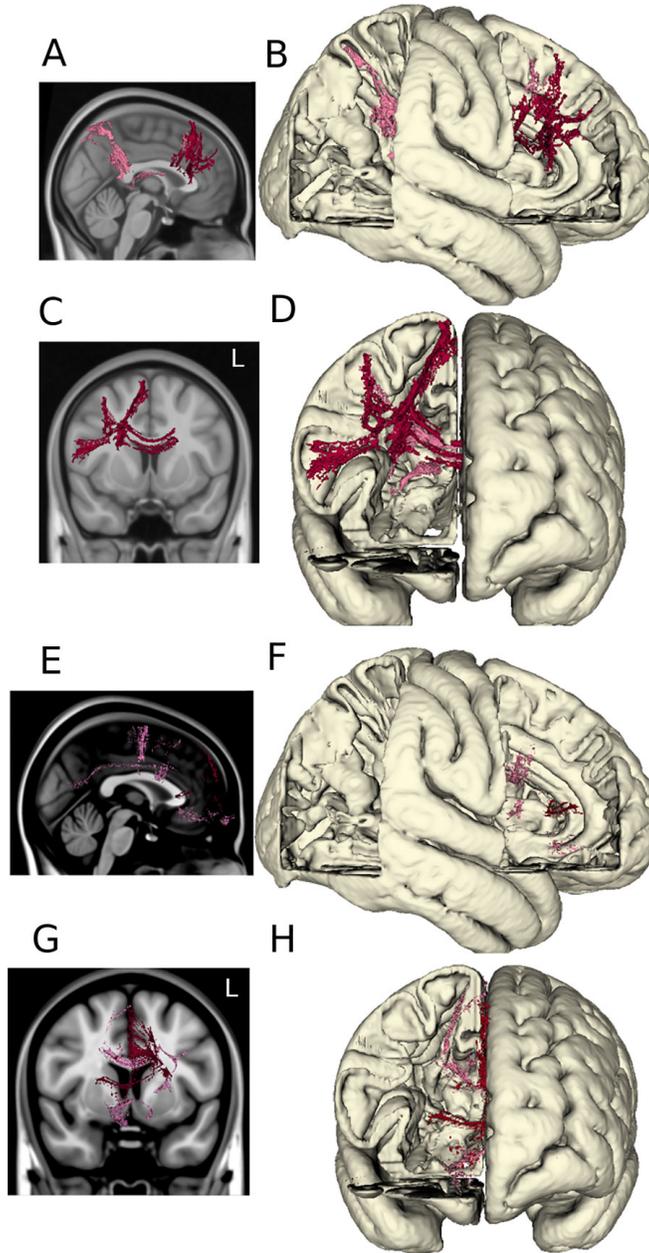


Figure A.1: DTI tractography. The probabilistic connectivity maps at optimal and optimal vs. suboptimal DBS settings are shown in sagittal and coronal projections with their respective 3D models. Panels (A–D): Significant voxels associated with optimal DBS settings (bilateral lower middle contacts, low amplitudes). Panels (E–H): Significant voxels comparing optimal vs. suboptimal DBS settings (bilateral lower middle contacts, high amplitudes).

Appendix B: Supplementary Materials for Chapter 3

B.1 Dual-stage two-process (DSTP) model implementation

B.1.1 Experiments 1 and 3

The DSTP model designed by Hübner and colleagues (2010) specifies two discrete stages of visual processing: 1) an early stage for identifying simple stimulus features and perceptual filtering, and 2) a late stage dedicated to processing the target. For reference, a diagram of the DSTP model is provided in *Figure 2* of the main manuscript. Stage 1 begins with two separate diffusion processes running in parallel, one representing the *stimulus selection phase* (which will be denoted “SS”) and the other representing the *response selection phase* (which will be denoted “RS1”). Evidence accumulation within each phase was implemented as a stochastic differential equation:

$$dx = \rho \frac{dt}{\Delta t} + \xi \sqrt{\frac{dt}{\Delta t}}$$

Here, x is evidence and ρ is the drift rate. The degree of noise in the accumulation process is represented by ξ , a driftless Wiener process distributed as $\xi \sim \mathcal{N}(0, 1)$. To approximate this continuous differential equation, we used the Euler method to discretize time, choosing a step size of $dt = 0.01$, modified by a time constant of $\Delta t = 0.1$. The drift rate for SS is a free parameter (ρ_{SS}) and the drift rate for RS1 is the sum of free parameters representing the strength of target and flanker stimuli, respectively:

$$\rho_{RS1} = \rho_{target} + \rho_{flankers}$$

where ρ_{target} is always positive and $\rho_{flankers}$ is negative in the incongruent task condition and positive in the congruent task condition. In SS and RS1, evidence accumulated through time until either process reached a decision threshold. Evidence accumulation processes were bounded between decision thresholds relevant to each phase (α_{SS} , α_{RS}) and 0. Starting points for each phase were determined from proportions z of the relevant threshold, such that

$$x_{RS1}(0) = z_{RS1}\alpha_{RS}$$

$$x_{SS}(0) = z_{SS}\alpha_{SS}$$

If x_{RS1} reached α_{RS1} or 0 before x_{SS} reached α_{SS} or 0, a response was made immediately with an RT equal to the sum of the duration of RS1 and non–decision time τ . In RS1, crossing the α_{RS} boundary meant the response corresponding to the target stimulus was selected, whereas crossing the 0 boundary meant the response corresponding to the flanker stimuli was selected. If x_{SS} reached α_{SS} or 0 before x_{RS1} reached α_{RS} or 0, a stimulus is selected for further processing in Stage 2. In SS, crossing the α_{SS} boundary indicated selection of the target for further processing, whereas crossing the 0 boundary indicated selection of the flankers for further processing. Response selection in Stage 2 (denoted “RS2”) is another diffusion process with drift rate ρ_{RS2} and threshold α_{RS} . The starting point $x_{RS2}(0)$ of RS2 was the value of x_{RS1} at time t when x_{SS} reached a decision boundary. ρ_{RS2} was negative when the stimulus was incongruent and x_{SS} crosses 0. As in Stage 1, crossing the α_{RS} boundary in Stage 2 meant the response corresponding to the target stimulus was selected, whereas crossing the 0 boundary meant that the response corresponding to the flanker stimuli was selected. The RT was equal to the sum of the durations of RS1 and RS2, and non decision time parameter τ . Free parameters and priors in our implementation of the DSTP model are provided in Table B.1.

Table B.1: Summary of free parameters and priors in the DSTP

| Parameter | Description | Prior |
|-------------------|-------------------------------|---------------------------------------|
| ρ_{target} | drift rate, RS1, target | $\mathcal{TN}(1.0, 4.0, 0.0, 10.0)$ |
| $\rho_{flankers}$ | drift rate, RS1, flankers | $\mathcal{TN}(1.0, 4.0, 0.0, 10.0)$ |
| ρ_{SS} | drift rate, SS | $\mathcal{TN}(0.0, 4.0, -10.0, 10.0)$ |
| ρ_{RS2} | drift rate, RS2 | $\mathcal{TN}(1.0, 4.0, 0.0, 10.0)$ |
| α_{RS} | decision threshold, RS1 & RS2 | $\mathcal{TN}(2.5, 5.0, 0.0, 20.0)$ |
| α_{SS} | decision threshold, SS | $\mathcal{TN}(2.5, 5.0, 0.0, 20.0)$ |
| z_{RS1} | starting point, RS1 | $\mathcal{TN}(0.5, 0.15, 0.0, 1.0)$ |
| z_{SS} | starting point, SS | $\mathcal{TN}(0.5, 0.15, 0.0, 1.0)$ |
| τ | non decision time | $\mathcal{U}(0.0, \min(RT))$ |

The model was able to capture basic effects such as a higher proportion of errors in the incongruent condition compared to congruent, and faster errors than correct responses in the incongruent condition. *Figure S1* illustrates the performance of the DSTP model via model-generated choice-RT distributions for each condition. The model performs similarly to the FFI time model, predicting more variability in RTs for correct responses in the incongruent condition than we observe in the data. As shown by conditional accuracy functions (CAFs) in *Figure B.2*, the DSTP model also does not predict slow errors in either the congruent or incongruent condition, similar to the FFI models in our investigation. Interestingly, the DSTP model is able to capture fast errors in the congruent condition, unlike the time-based models in our investigation. As discussed in the main text, the poor performance of the DSTP model in comparison to the conflict-based alternatives appears to be due in part to its complexity.

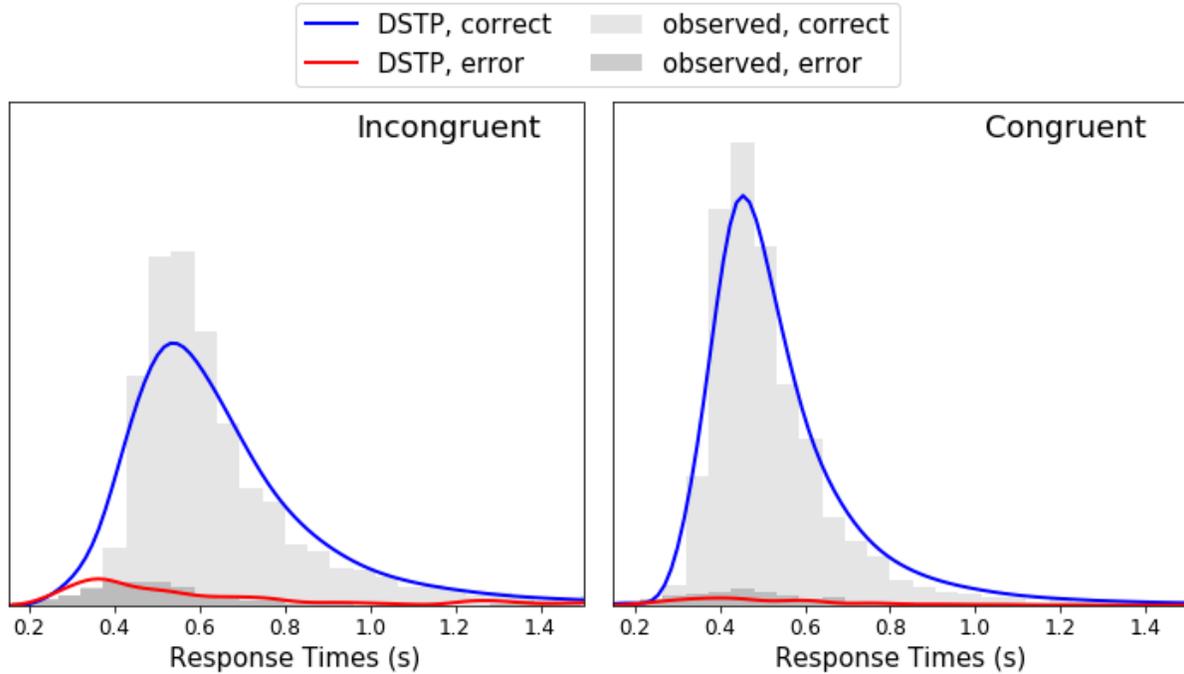


Figure B.1: Observed and DSTP model-generated choice-RT distributions. Observed RT distributions for correct (light gray histograms) and incorrect (dark gray histograms) responses were averaged across participants. Models were simulated 10,000 times for each condition, using each participant's best-fitting parameters. Lines show average model-generated distributions across participants.

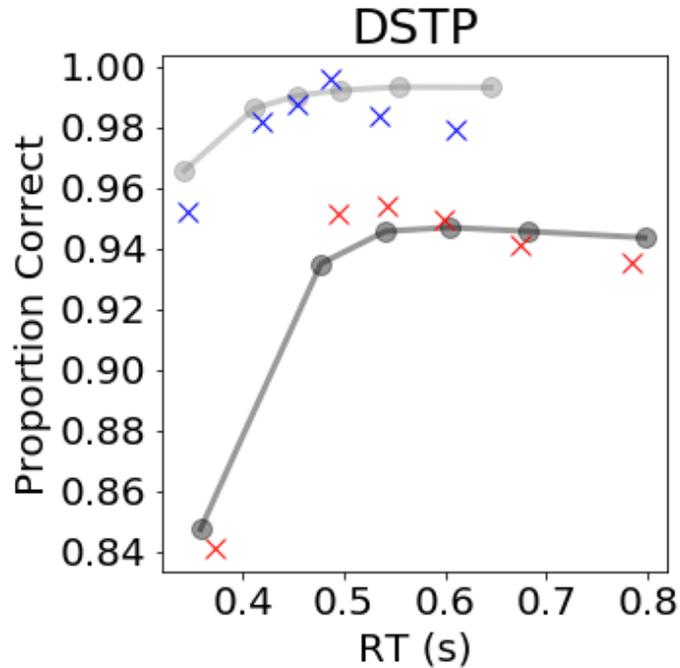


Figure B.2: Observed and DSTP-predicted CAFs for congruent and incongruent trials. Data from each subject were sorted according to RT within 6 equally-spaced percentile bins. Performance and minimum RT for each bin were averaged across participants in the congruent (blue Xs) and incongruent (red Xs) conditions. After generating 1,000 choice-RT pairs from each subject's best-fitting parameters within each model, the same procedure was used to calculate CAFs for each model (gray lines).

B.1.2 Experiment 2

Experiment 2, designed and administered by Servant et al. (2014), required participants to indicate the color of a target circle amid flanker circles of a congruent or incongruent color. The color saturation of center target circles varied from trial to trial within six conditions (degrees of suprathreshold saturation levels: 15%, 25%, 35%, 45%, 60% and 80%), while the color saturation of flanker circles was held constant at 80%. An example of the stimuli used in Experiment 2 is provided in *Figure 13* of the main manuscript. To fit the DSTP model, we

needed to make adjustments to accommodate the target color saturation manipulation. When modifying our SSP variants to fit data from Experiment 2, we replaced the perceptual input strength parameter p with six points along a monotonically increasing sigmoid function. We took a similar approach to modifying the DSTP. We made the assumption that the target color saturation manipulation would affect both RS1 and SS phases of Stage 1, with RS1 representing automatic feature-driven attentional processes and SS representing a more controlled mode of selecting a stimulus for further processing. Because RS2 in Stage 2 represents decision processes after the target has already been identified in the SS phase, specific perceptual features of the target like color saturation should not have an effect on ρ_{RS2} . We therefore implemented sigmoidal functions to calculate drift rates in both RS1 and SS. Because the color saturation of targets varied between trials while the color saturation of flanker stimuli was held constant, we specified a vector k such that

$$k = [\rho_{target(0.15)}, \rho_{target(0.25)}, \rho_{target(0.35)}, \rho_{target(0.45)}, \rho_{target(0.60)}, \rho_{target(0.80)}]$$

was calculated via a sigmoidal function

$$k_i = d_{RS1} + \frac{a_{RS1} - d_{RS1}}{1 + e^{-c_{RS1}(h_i - b_{RS1})}}$$

where $h = [0.15, 0.25, 0.35, 0.45, 0.60, 0.80]$ and a_{RS1} , b_{RS1} , and c_{RS1} were free parameters.

d_{RS1} , representing the floor value of the sigmoidal function, was fixed to 0 since values of ρ_{target} were constrained to be positive in the original model. ρ_{RS1} was then calculated with the equation

$$\rho_{RS1} = \rho_{target(i)} + \rho_{flankers}$$

for each target color saturation condition i , where $\rho_{flankers}$ was negative on incongruent trials.

Similarly for the SS phase, we specified a vector j such that

$$j = [\rho_{SS(0.15)}, \rho_{SS(0.25)}, \rho_{SS(0.35)}, \rho_{SS(0.45)}, \rho_{SS(0.60)}, \rho_{SS(0.80)}]$$

was calculated via a sigmoidal function

$$j_i = d_{SS} + \frac{a_{SS} - d_{SS}}{1 + e^{-c_{SS}(h_i - b_{SS})}}$$

Where a_{SS} , b_{SS} , c_{SS} , and d_{SS} were free parameters. Here, d_{SS} was free to allow drift rates in the SS phase to take on negative values. Examples of sigmoidal functions calculated from various values of a , b , and c are shown in *Figure 14* of the main manuscript. Priors for free parameters governing the sigmoidal functions for phases RS1 and SS are provided in Table B.2.

Table B.2: Summary of free parameters and priors in the DSTP added for Experiment 2

| Parameter | Prior |
|-----------|--|
| a_{RS1} | $\mathcal{TN}(1.0, 10.0, 0.0, 20.0)$ |
| b_{RS1} | $\mathcal{U}(-1.0, 1.0)$ |
| c_{RS1} | $\mathcal{TN}(4.0, 10.0, 0.0, 30.0)$ |
| a_{SS} | $\mathcal{TN}(1.0, 10.0, -20.0, 20.0)$ |
| b_{SS} | $\mathcal{U}(-1.0, 1.0)$ |
| c_{SS} | $\mathcal{TN}(4.0, 10.0, 0.0, 30.0)$ |
| d_{SS} | $\mathcal{TN}(0.0, 10.0, -20.0, 20.0)$ |

B.2 Additional models fit to data from Experiment 1

B.2.1 LCA control, fixed κ and β

The LCA control model that was described in the main manuscript included free parameters for both leak (κ) and lateral inhibition (β). To test if either κ or β was independently driving the model's success at fitting the data from Experiment 1, we additionally fit variants with free κ and fixed β ($\beta = 0$), and free β and fixed κ ($\kappa = 0$). *Figure S5* is a modified version of *Figure 6*, which illustrates BPIC values for each model, mean-centered within-subject. Our results show that the full LCA control model with free κ and β outperform the fixed-parameter variants. By

calculating differences in BPIC values between the fixed κ and fixed β variants across subjects, we found that the model with fixed κ fit better than the model with fixed β on average (15 wins for model with fixed κ compared to 11 wins for the model with fixed β). Compared to the full LCA-control model with free leak and lateral inhibition terms, however, neither model fit the data as well (14 wins for full model, 6 wins for model with fixed κ , 6 wins for model with fixed β). Our results indicate that both leak and lateral inhibition are necessary for fitting the data across subjects within our control framework.

B.2.2 LCA exponential time

Although a linear definition of time was adopted within the original SSP, we developed a model variant in which time took the form of an exponential function that increased up to an asymptote. Because cognitive control in the LCA control model tends to increase throughout a trial and taper off toward the end, we wanted to see if a sublinear functional form for the signal guiding the attentional spotlight was sufficient for superior fits, independent of the cognitive control mechanism itself. As such, the ‘‘LCA exponential time’’ model was designed to be similar to LCA control. Similar to the equations for cognitive control (c) provided in *Section 2.2.2*, the spotlight width in LCA exponential time is given by:

$$sd_a = sd_0 - r_d t_{exp}$$

where the change in t_{exp} is given by:

$$dt_{exp} = (\delta - t_{exp}) \frac{dt}{\Delta t}.$$

As in the conflict models, the δ prior was distributed as $\delta \sim \mathcal{TN}(2.5, 10, 0, 30)$. The LCA exponential time model contained the same 8 free parameters as LCA control. *Figure S3* illustrates the contrasting functional forms of within-trial signals guiding the spotlight across

linear time, control, and exponential time model variants. The functional form of the exponential time signal monotonically increases up to an asymptote, represented by the free parameter δ . For similar implementations of an exponential function calculated as the distance from an asymptote, see Steyvers, Hawkins, Karayanidis, & Brown, 2019 and Turner, 2019.

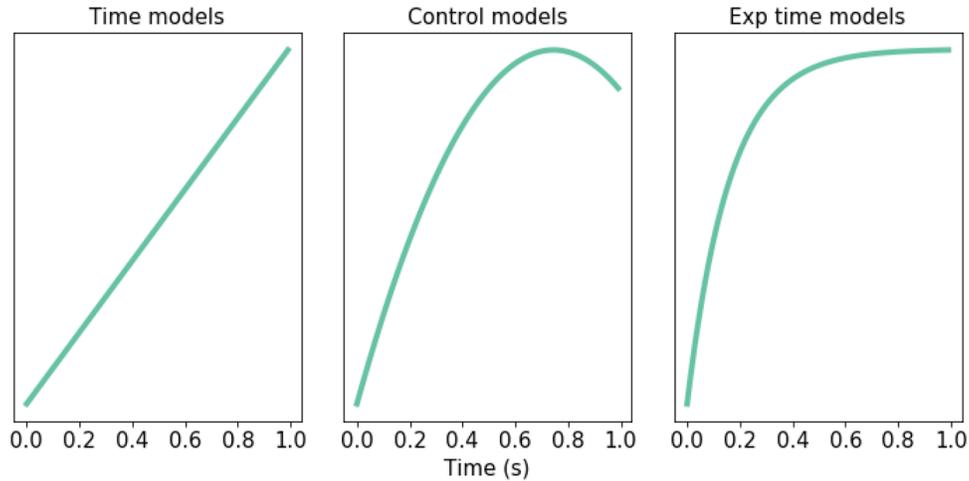


Figure B.3: Linear time-, control-, and exponential time-based signals to attention. Noiseless simulations of each signal type are shown for a single trial.

We fit the model to data from Experiment 1 using the procedures described in Section 3.1.1.4. Model fits were verified by generating data from each participant's best-fitting parameters, and comparing model-generated choice-RT distributions to the participant's observed data. Mean distributions of observed and model-generated RTs for correct and incorrect responses in both task conditions are shown in Figure B.4.

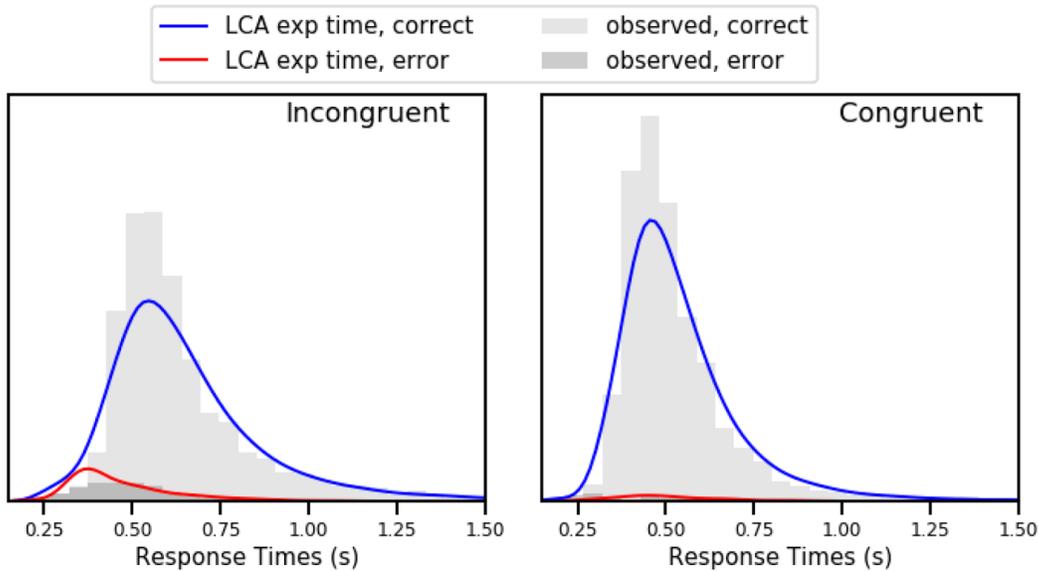


Figure B.4: Observed and LCA exponential time model-generated choice-RT distributions.

Observed RT distributions for correct (light gray histograms) and incorrect (dark gray histograms) responses were averaged across participants. Models were simulated 10,000 times for each condition, using each participant's best-fitting parameters. Lines show average model-generated distributions across participants.

Despite providing decent qualitative fits to data, the LCA exponential time model did not fare well when quantitatively compared to the other models in our investigation. As shown by the heatmap of within-subject mean-centered BPIC values in Figure B.5, LCA exponential time was the worst-performing model across all alternatives. Although there remain an infinite number of ways we could have defined time, these results indicate that the cognitive control mechanism in the LCA control model is accounting for aspects of the data that a similar, time-based mechanism could not.

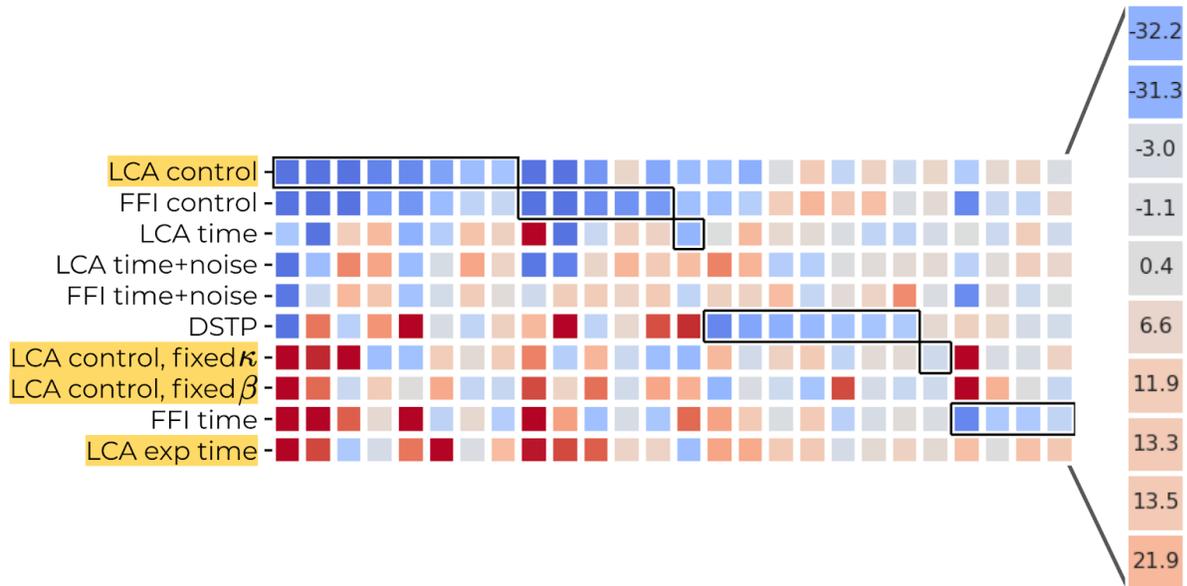


Figure B.5: Heat map of BPIC values, mean-centered within-subject for Experiment 1. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right.

B.3 Additional results from Experiment 3

A comparison of model fits to the behavioral data collected in Experiment 3 is shown as a heatmap in Figure B.6. BPIC values were mean-centered within-subject, and lower values indicate better fits. When considering mean values across subjects, we observed similar results in Experiment 3 as in Experiment 1, such that the two control-based models outperformed the alternatives.

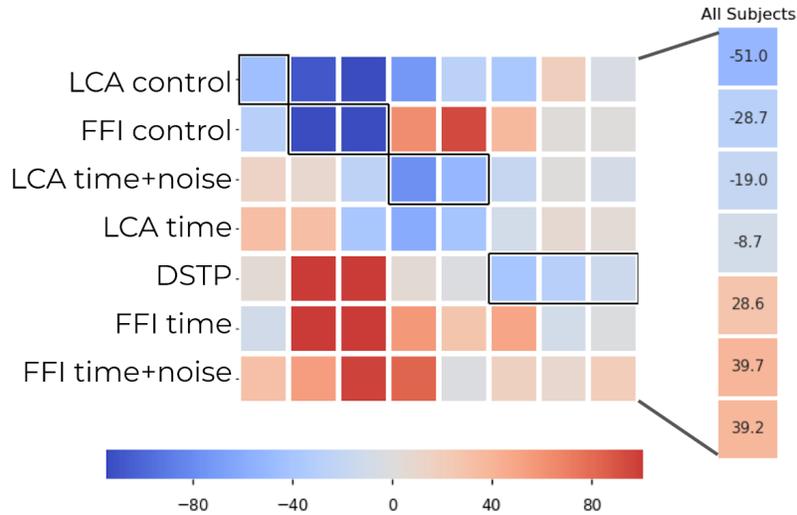


Figure B.6: Heat map of BPIC values, mean-centered within-subject for Experiment 3. Each column corresponds to a subject. Lower BPIC values (blue hues) indicate better model fits. The winning model for each subject is outlined in black. Average mean-centered values across subjects are shown in the panel to the right.

In our model-based EEG analysis of data from a standard flanker task administered in Experiment 3, we calculated correlations between the signals controlling the width of the attentional spotlight (e.g. time, time+noise, or cognitive control) and EEG voltage during the decision. Correlation maps for each individual model are shown in Figure 3.19 of the main manuscript. Topographic plots in Figure B.7 show that increased correlations between EEG voltage and attention modulation in LCA control, relative to the predictions of the other models, are widespread across the scalp. None of the other difference maps reveal significant results.

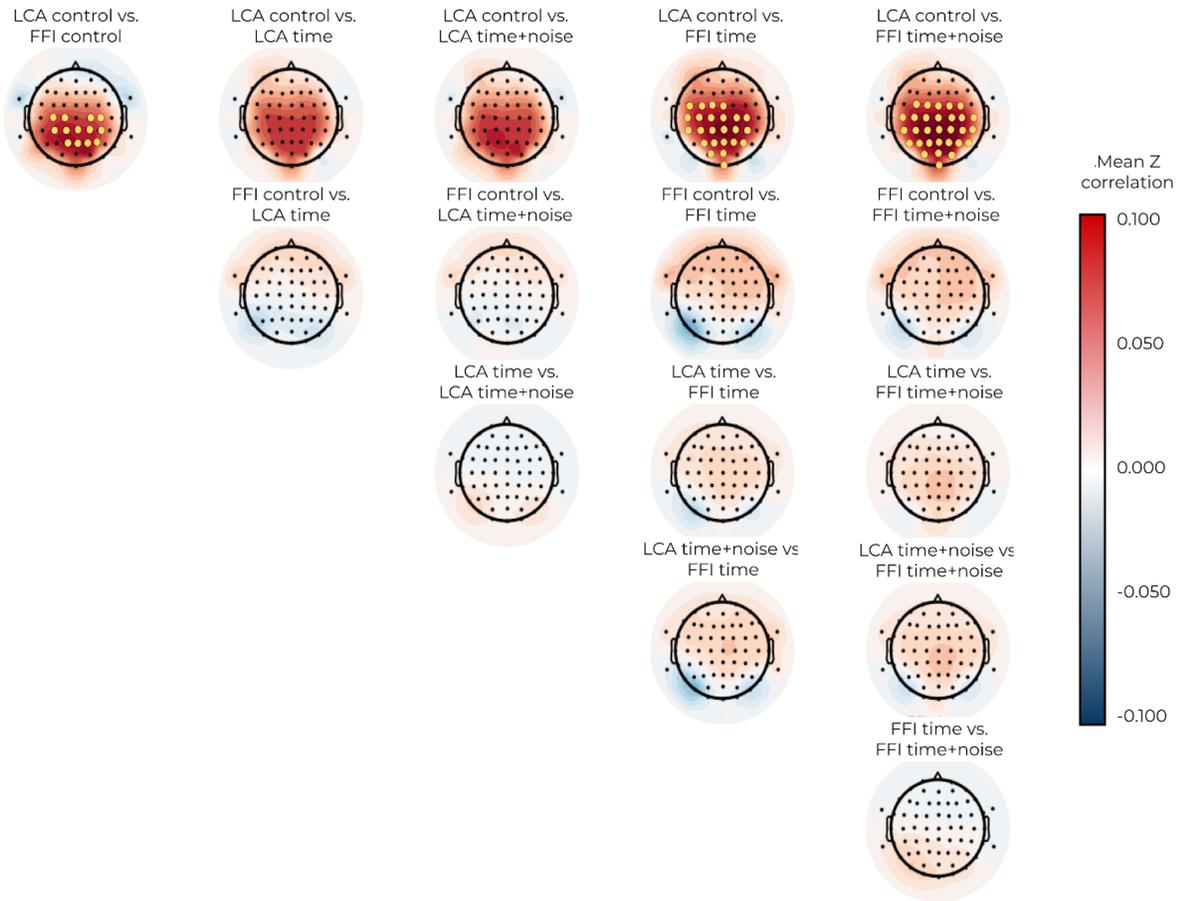


Figure B.7: Mean Z correlation difference maps for observed EEG data and model-generated attention modulation signals. After calculating Z correlation values for each model and each electrode, we calculated the pairwise difference topographic maps for each possible pair of models. P values were calculated for each model comparison and electrode using a 1-sample t -test. Significant correlation differences were identified using a Benjamini-Hochberg correction for multiple comparisons, indicated by yellow points.