

NEUROBIOLOGICAL FOUNDATIONS OF STABILITY AND FLEXIBILITY

Nathan Tardiff

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Sharon L. Thompson-Schill, Professor of Psychology

Graduate Group Chairperson

John Trueswell, Professor of Psychology

Dissertation Committee

Michael L. Platt, Professor of Neuroscience, Psychology, and Marketing

Joseph Kable, Professor of Psychology

Lyle Ungar, Professor of Computer and Information Science

ACKNOWLEDGMENTS

They say it takes a village to raise a doctoral student, and my case is no exception. First, I am grateful to my advisor, Sharon Thompson-Schill, for giving me the opportunity to pursue my graduate studies in her lab, and for giving me freedom and support in developing the work presented here. Her advice and guidance related to this work and to the wider world of academia have been invaluable, and I feel incredibly fortunate. I also want to thank my committee, Michael Platt, Joe Kable, and Lyle Ungar, whose keen feedback has also helped shaped this work and my growth as a researcher.

I also want to thank my fellow members of the Thompson-Schill lab past and present, who have been a tremendous source of academic and scientific support, as well as support in the emotional labor of graduate school. I especially want to thank Kat Graves, who contributed immensely to this work through tireless efforts in data collection, technical assistance, and pun-making. One could not ask for a better friend and collaborator. Special appreciation also goes out to Sarah Solomon, Lisa Musz, Zuzanna Balewski, John Medaglia, Lizz Karuza, Anna Leshinsky, and Marcelo Mattar.

For the work presented in Chapter 2, thanks are due to Roy Hamilton, Olu Faseyitan, and the LCNS lab for assistance with tDCS current modeling and stimulation parameters and to Hetty Rodriguez, Jennifer-Marie Rosado, and the Penn Molecular Profiling Facility for genotyping services. I also thank Brianna Owairu for help with data collection. For Chapter 3, I'd like to acknowledge helpful feedback from Josh Gold, Chris Glaze, and other members of the Gold lab, as well as Dani Bassett, Jeni Stiso, Ursula Tooley, and

others in the Complex Systems Lab. I'd also like to thank Mark Elliott, Karthik Prabhakaran, and Andrew Bock for technical assistance. Another shout-out is also due to Kat and Zuzanna for scanner and eye tracking wizardry. Additionally, all of the work in this thesis was supported by a National Defense Science and Engineering Graduate Fellowship.

I also want to give my deepest appreciation to my family, Laura and Asher. Living with a graduate student isn't always easy, and I am lucky that they have supported me, put up with me, and kept me grounded throughout this process. I also am thankful for the support of my parents and in-laws.

Finally, I'd like to thank Debbie Zaitchik, whose friendship and belief in me set me on this path.

The findings from Chapter 2 have been published as: Tardiff, N., Graves, K. N., & Thompson-Schill, S. L. (2018). The role of frontostriatal systems in instructed reinforcement learning: Evidence from genetic and experimentally-induced variation. *Frontiers in Human Neuroscience*, 12, 472. Copyright retained by the authors.

ABSTRACT

NEUROBIOLOGICAL FOUNDATIONS OF STABILITY AND FLEXIBILITY

Nathan Tardiff

Sharon L. Thompson-Schill

In order to adapt to changing and uncertain environments, humans and other organisms must balance stability and flexibility in learning and behavior. Stability is necessary to learn environmental regularities and support ongoing behavior, while flexibility is necessary when beliefs need to be revised or behavioral strategies need to be changed. Adjusting the balance between stability and flexibility must often be based on endogenously generated decisions that are informed by information from the environment but not dictated explicitly. This dissertation examines the neurobiological bases of such endogenous flexibility, focusing in particular on the role of prefrontally-mediated cognitive control processes and the neuromodulatory actions of dopaminergic and noradrenergic systems. In the first study (Chapter 2), we examined the role of frontostriatal circuits in instructed reinforcement learning. In this paradigm, inaccurate instructions are given prior to trial-and-error learning, leading to bias in learning and choice. Abandoning the instructions thus necessitates flexibility. We utilized transcranial direct current stimulation over dorsolateral prefrontal cortex to try to establish a causal role for this area in this bias. We also assayed two genes, the COMT Val158Met genetic polymorphism and the DAT1/SLC6A3 variable number tandem repeat, which affect prefrontal and striatal dopamine, respectively. The results support the role of prefrontal cortex in biasing learning, and provide further evidence that individual differences in the

balance between prefrontal and striatal dopamine may be particularly important in the tradeoff between stability and flexibility. In the second study (Chapter 3), we assess the neurobiological mechanisms of stability and flexibility in the context of exploration, utilizing fMRI to examine dynamic changes in functional brain networks associated with exploratory choices. We then relate those changes to changes in norepinephrine activity, as measured indirectly via pupil diameter. We find tentative support for the hypothesis that increased norepinephrine activity around exploration facilitates the reorganization of functional brain networks, potentially providing a substrate for flexible exploratory states. Together, this work provides further support for the framework that stability and flexibility entail both costs and benefits, and that optimizing the balance between the two involves interactions of learning and cognitive control systems under the influence of catecholamines.

TABLE OF CONTENTS

ACKNOWLEDGMENTS II

ABSTRACT.....IV

LIST OF TABLES VII

LIST OF ILLUSTRATIONSVIII

I. INTRODUCTION..... 1

II. THE ROLE OF FRONTOSTRIATAL SYSTEMS IN INSTRUCTED REINFORCEMENT LEARNING: EVIDENCE FROM GENETIC AND EXPERIMENTALLY-INDUCED VARIATION..... 10

1. Introduction 10

2. Methods..... 14

3. Results 23

4. Discussion 46

III. THE MODULATION OF BRAIN NETWORK INTEGRATION AND AROUSAL DURING EXPLORATION..... 58

1. Introduction 58

2. Methods..... 62

3. Results 79

4. Discussion 100

IV: GENERAL DISCUSSION..... 114

APPENDIX A: SUPPLEMENTARY MATERIAL FOR CHAPTER 2 122

APPENDIX B: SUPPLEMENTARY MATERIAL FOR CHAPTER 3 142

BIBLIOGRAPHY 149

LIST OF TABLES

Table 2.1.	15
Table 2.2.	18
Table 2.3.	39
Table 2.4.	45

LIST OF ILLUSTRATIONS

Figure 2.1	28
Figure 2.2	30
Figure 2.3	41
Figure 3.1	63
Figure 3.2	80
Figure 3.3	82
Figure 3.4	84
Figure 3.5	87
Figure 3.6	89
Figure 3.7	90
Figure 3.8	94
Figure 3.9	96
Figure 3.10	99

I. INTRODUCTION

Behavioral flexibility is crucial to survival in changing and uncertain environments. At any given time, an organism must decide whether to continue pursuing the current behavioral policy, thereby maintaining stability, or flexibly abandon that policy in favor of alternative and potentially more beneficial goals and courses of action. A squirrel foraging for acorns must decide when to abandon the current tree in favor of other trees with more abundant acorns. If a tree dies or the acorn yield is particularly poor, the squirrel must stop relying on that resource and find and remember other sources of food, such as human refuse. In humans, such flexibility extends beyond basic survival decisions and extends across timescales. In cities, we must be flexible navigators, adapting to continual change brought about by traffic, potholes, and construction. We demonstrate flexibility in our preferences, as a young child who enjoys a superhero show one month only to switch to a different show the next month. We must be flexible at work and in our careers, deciding when to abandon one project or job in favor of a better path. At the same time, humans display remarkable stability. For example, we can focus on long-term goals like obtaining a degree, at the expense of short-term payoffs.

As these examples demonstrate, flexibility broadly construed involves the coordination of learning and decision-making capacities; it often depends on balancing reliance on prior knowledge with learning new knowledge that may override prior beliefs, as well as balancing exploiting a resource with exploring to find other, potentially better resources. Both stability and flexibility entail costs and benefits (Blackwell, Chatham, Wiseheart, & Munakata, 2014; Cools & D'Esposito, 2009; Friedman & Miyake, 2017; Gopnik,

Griffiths, & Lucas, 2015; Herd et al., 2014; Nassar, Wilson, Heasley, & Gold, 2010). Knowledge or behavior that is too stable results in rigidity and inflexibility, as exemplified by over-trained animals who continue to level press for food, wasting time and energy that could be better spent elsewhere (Niv, 2009). Conversely, too much flexibility is characterized by learning that is too influenced by recent experience, leading the organism to miss important environmental regularities, or by behavior that is not strongly organized by internal goals, leading to distraction and disinhibition (Cools & D'Esposito, 2009; Nassar et al., 2010).

The motivating questions addressed by this thesis concern the neurobiological substrates of stability and flexibility, focusing on mechanisms by which the balance between these capacities are adjusted, including dynamic adjustment within individuals as well as differences across individuals. Below, we briefly review work that has begun to elucidate some of the computational and neurobiological mechanisms of flexibility and adaptive behavior, which will be shown to rely on prefrontally-mediated cognitive control and learning processes, both of which are powerfully influenced by the neuromodulatory actions of dopaminergic and noradrenergic systems. We will then describe how the work in this thesis extends these findings, focusing on neural substrates dictating the balance between stability and flexibility.

Within human cognitive neuroscience and neuropsychology, the study of flexibility in thought and action has uncovered a set of mechanisms known as executive function or cognitive control. The primary components of executive function most often cited are working memory, shifting, and inhibition (Diamond, 2013; Friedman & Miyake, 2017).

The exertion of various combinations of these faculties allows people to plan and execute actions extended over time, flexibly switch between different tasks or rules, and override prepotent or overlearned responses, all in accordance with internal goals. These capabilities are dependent on the integrity of the prefrontal cortex (Diamond, 2013). In particular, the prefrontal cortex (PFC) is thought to provide top-down signals that bias processing in other brain areas, thereby facilitating processing that is aligned with current goals while suppressing processing that conflicts with these goals (Miller & Cohen, 2001).

While these abilities are no doubt necessary for flexible behavior, as is evident in the perseverative behavior of young children, in whom the prefrontal cortex is still developing (Munakata, Snyder, & Chatham, 2012; Zelazo et al., 2003), the tasks measuring these abilities in human subjects generally involve following explicitly provided rules in deterministic environments with restricted opportunities for learning. Even in switching tasks meant to tap cognitive flexibility, the need to switch is usually explicitly cued. Though such exogenously cued switching no doubt taps important aspects of cognitive flexibility, it fails to capture the endogenous flexibility people must deploy in everyday environments that do not contain explicit cues on when to switch and which rules to switch to.

Of course, most organisms do not have the option to follow explicit instructions. Within neuroscience and computer science, reinforcement learning (RL) has been a dominant computational framework for understanding how organisms can learn to adapt their behavior in order to optimize reward (Dolan & Dayan, 2013; Niv, 2009; Sutton & Barto,

2018). A key insight from this literature is that by computing a reward prediction error (RPE)—the difference between the reward expected and the reward earned—organisms can learn to incrementally update their behavioral policies to ultimately maximize their reward (Schultz, Dayan, & Montague, 1997). However, despite this ability to learn to optimize behavior without explicit instructions, standard models of RL assume stationary environments and are inflexible in the face of change, needing to slowly learn new associations in order to alter behavior (Pearson, Heilbronner, Barack, Hayden, & Platt, 2011).

In sum, we argue that cognitive control and RL are both individually necessary but not sufficient for explaining the full range of adaptive behavior. In particular, they do not capture the capacity for *endogenous flexibility*, the ability to adapt cognition and behavior in a self-directed manner in order to meet the demands of uncertain and/or changing environments. This type of flexibility is distinct from the mere ability to switch behavior (avoid perseverating) in that it must be enacted without explicit environmental cues such as learned stimulus-response associations or explicit verbal instructions.

More recent efforts have begun to characterize the computational and neural underpinnings of learning and control in dynamic, uncertain, or novel environments. This work has begun to point toward a synthesis of learning and cognitive control, demonstrating that they are interdependent in promoting endogenously flexibility (e.g., Cohen, McClure, & Yu, 2007; Collins & Koechlin, 2012; Pearson et al., 2011; Shenhav, Botvinick, & Cohen, 2013). For example, it appears that people can learn the statistics of a volatile environment through adaptive RL or Bayesian learning in order to adjust the

level of control they bring to a task (Jiang, Beck, Heller, & Egner, 2015). Similarly, research in reinforcement learning has demonstrated how flexible, model-based reinforcement learning, which utilizes goal-directed planning to overcome the limitations of standard RL, appears to rely on neural and cognitive processes that overlap with those of cognitive control (Doll, Bath, Daw, & Frank, 2016; Otto, Gershman, Markman, & Daw, 2013a; Otto, Raio, Chiang, Phelps, & Daw, 2013b; Otto, Skatova, Madlon-Kay, & Daw, 2015; Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013).

Of particular importance to the present work, the neuromodulatory actions of dopamine (DA) and norepinephrine (NE) are thought to play a key role in the cognitive control and adaptive learning processes necessary for adjusting the balance between stability and flexibility. Phasic responses of midbrain dopamine neurons have been shown to signal reward prediction error (Schultz et al., 1997), which is important for updating the expected value of both overt actions and the internal action of updating working memory in corticostriatal circuits (Niv & Schoenbaum, 2008; O'Reilly & Frank, 2006; Schultz et al., 1997). Surprise signals derived from unsigned RPE can be conveyed to brain areas such as the anterior cingulate cortex to drive adjustments in behavior, including exploration (Hayden, Heilbronner, Pearson, & Platt, 2011). NE—a key modulator of physiological arousal that is released by neurons in the locus coeruleus (LC)—has been ascribed a number of computational roles, including signaling uncertainty and the probability of an environmental change, quantities that can be used to dynamically adjust learning rates (Nassar et al., 2012), effectively changing the balance between bottom-up and top-down information (Yu & Dayan, 2005). NE has also been suggested to mediate

the balance between exploration and exploitation (Aston-Jones & Cohen, 2005) or to reorganize functional brain networks for different behavioral demands (Bouret & Sara, 2005).

Both DA and NE strongly affect prefrontal circuits, with optimal prefrontal functioning occurring at moderate levels of each (Arnsten, 2011). For example, prefrontal DA levels related to both genotypic variation and pharmacological manipulations have been associated with differences in the stability of representations in prefrontal working memory (Cools & D'Esposito, 2009). The Met allele of the Val158Met COMT genetic polymorphism is associated with higher baseline levels of prefrontal DA and better maintenance of information in working memory, while the Val allele is associated with lower prefrontal DA and worse maintenance of information in working memory, but more flexible updating of working memory (Cools & D'Esposito, 2009). This advantage in flexibility for the Val allele has also been shown to extend to reinforcement learning paradigms, in which Val homozygotes more flexibly adapt to reversals (Krugel, Biele, Mohr, Li, & Heekeren, 2009). Computational models of working memory suggest that this COMT-mediated stability-flexibility tradeoff is a necessary consequence of a working memory system that must be both robust to interference and able to be rapidly updated as the situation demands (Durstewitz & Seamans, 2008; O'Reilly & Frank, 2006). DA-mediated changes in the balance between stability and flexibility are explained by changes in the attractor dynamics of prefrontal networks, which do not allow for the simultaneous coexistence of flexibility and stability in one state (Durstewitz & Seamans, 2008), or alternatively by striatal mechanisms that gate access to PFC (O'Reilly & Frank,

2006).

This thesis extends the study of the modulation of stability and flexibility in situations requiring endogenously initiated changes in control state. Chapter 2 addresses the computational and neurobiological substrates of endogenous flexibility in the context of instructed reinforcement learning, a class of paradigms in which verbal instructions given prior to learning influence learning and choice. Both neuroimaging and neuogenetics have suggested a role for frontostriatal circuits in biasing instructed RL (Doll, Hutchison, & Frank, 2011; Fouragnan et al., 2013; Li, Delgado, & Phelps, 2011). In particular, the COMT Val158Met genetic polymorphism discussed above is associated with the degree of instructional bias (Doll et al., 2011). Chapter 2 addressed three main goals. First, we sought to replicate the effect of COMT on instructed RL, providing further evidence for the role of PFC-mediated top-down control in biasing RL. Second, we aimed to expand the understanding of the impact of dopaminergic genes on instructed RL by examining the effect of the DAT1/SLC6A3 variable number tandem repeat (VNTR), which affects striatal DA reuptake (Faraone, Spencer, Madras, Zhang-James, & Biederman, 2014). Striatal DA levels have previously been linked to cognitive flexibility (Cools & D'Esposito, 2009), making DAT1 a plausible but as yet unassessed modulator of instructed RL. Finally, we hoped to establish a causal link between PFC and instructional bias by directly modulating PFC via transcranial direct current stimulation (tDCS). In particular, we hypothesized that anodal stimulation would increase bias and that cathodal stimulation would decrease it, though this latter hypothesis was more tentative given the unreliability of cathodal stimulation in cognitive tasks (Jacobson, Koslowsky, & Lavidor,

2012). Together, this chapter aims to provide further evidence for the role of frontostrially-mediated cognitive control processes in biasing RL, highlighting individual differences in dopaminergic function associated with differences in the ability to flexibly adapt behavior. The results are interpreted within a framework that argues that top-down control can incur both costs and benefits, depending on its fit to the task (Chrysikou, Weber, & Thompson-Schill, 2014).

While DA is more associated with regulating stability and flexibility in frontostriatal circuits, NE is thought to have widespread effects throughout the cortex (Berridge & Waterhouse, 2003b), making it well-situated to exert global influences on stability and flexibility. As noted above, theories of locus coeruleus-norepinephrine (LC-NE) function have ascribed it a key role in adjusting this balance, potentially by facilitating the reconfiguration of brain networks (Bouret & Sara, 2005). In the human neuroimaging literature, a number of recent studies utilizing pupil diameter as an indirect marker of LC activity or using pharmacological manipulation of NE have found evidence in favor of NE's role in reshaping functional brain networks (e.g., Eldar, Cohen, & Niv, 2013; Shine et al., 2016; van den Brink et al., 2016). Notably, some of these studies suggest that NE levels can alter the balance of functional coupling, or integration, between different brain networks.

Utilizing network neuroscience methods and pupillometry, Chapter 3 probes the relationship between brain network dynamics and LC-NE system activity in the context of switching between exploration (flexibility) and exploitation (stability). This study aimed to address multiple shortcomings of the prior literature. First, to date the

relationship between NE and functional connectivity has not been assessed within the context of a task with an established relationship between NE-associated arousal and behavior, with prior studies relying on incidental variations in arousal or pharmacological manipulation. Second, most prior work has relied on static brain networks constructed over long periods of time, making it difficult to establish whether connectivity changes dynamically track changes in NE. In accomplishing this goal, we sought to make methodological advances by demonstrating the ability to detect changes in brain network integration at a much finer temporal scale than is generally examined. In sum, by more tightly linking brain network dynamics, LC-NE associated activity, and exploratory choice, this chapter is intended to further our understanding of NE's role in mediating between stability and flexibility.

Chapter 4 synthesizes the findings of Chapters 2 and 3 and suggests possible directions for future work.

II. THE ROLE OF FRONTOSTRIATAL SYSTEMS IN INSTRUCTED REINFORCEMENT LEARNING: EVIDENCE FROM GENETIC AND EXPERIMENTALLY-INDUCED VARIATION

1. Introduction

Successful learning and decision-making require a balance between exploiting prior information and learning from new experiences that may contradict it. One pervasive source of prior information in humans is instruction from others. Such instruction has clear benefits on both ontogenetic and historical timescales, allowing children to rapidly learn about the world and allowing culture and technology to develop and evolve (Tomasello, 1999). On an individual level, advice and information received from friends, professionals, and the media shape our view of the world and our choices.

The alternative to learning from advice and instruction is learning from direct experience of the world. One well-characterized method of learning from experience is reinforcement learning (RL), in which actions are selected so as to maximize reward (see Dolan & Dayan, 2013 and Niv, 2009 for reviews). Recent work exploring the effects of instruction on RL has found that accurate advice can significantly improve performance (Biele, Rieskamp, & Gonzalez, 2009; Doll et al., 2011). Yet such instruction is often detrimental when it is inaccurate. A potential consequence of inaccurate instruction and, more generally, inaccurate prior information, is *confirmation bias*, whereby data that are consistent with a prior hypothesis are sought, attended to, or valued over disconfirming data, which are neglected, filtered, or devalued (Nickerson, 1998). Confirmation bias is

thought to be pervasive in human reasoning, affecting children and adults' scientific reasoning as well as that of professional scientists (Hergovich, Schott, & Burger, 2010; Kuhn, 1989; MacCoun, 1998; Mahoney, 1977).

Biases have been induced in both social and nonsocial RL tasks utilizing various methods of information delivery. Information indicative of the moral character of computerized partners in a repeated trust game biases share decisions to “good” and “bad” partners despite identical behavior by the computer (Delgado, Frank, & Phelps, 2005; Fareri, Chang, & Delgado, 2012). Poor advice provided by fellow subjects impairs performance on the Iowa Gambling Task (Biele et al., 2009; Biele, Rieskamp, Krugel, & Heekeren, 2011). Finally, in an RL task in which subjects learn to discriminate among pairs of probabilistically rewarded symbols, subjects instructed that a particular symbol is desirable persist in choosing that symbol more than would be expected given negative feedback, selecting it more frequently than symbols rewarded at an equal rate (Doll et al., 2011; Doll, Jacobs, Sanfey, & Frank, 2009; Doll et al., 2014; Staudinger & Büchel, 2013). In sum, instructional biases appear to be persistent, and they are only partially ameliorated by feedback.

The neural substrates of instructed learning are still emerging, though as in uninstructed RL, frontostriatal areas are commonly implicated (Doll et al., 2009; Wolfensteller & Ruge, 2012). Neuroimaging has supported a role for prefrontal cortex (PFC) in representing instructions or prior information (Fouragnan et al., 2013; Li et al., 2011), with activity in instructed conditions found in dorsolateral PFC (DLPFC) and medial PFC. Connectivity analyses further support a role for PFC, reporting increased functional

connectivity between frontal and striatal regions during instructed/prior knowledge conditions, consistent with top-down influence on striatal reward prediction errors (Fouragnan et al., 2013; Li et al., 2011).

Evidence of PFC altering striatal learning comports well with accounts of PFC-mediated cognitive control biasing or filtering information in other brain regions. Such top-down modulation focuses information processing on task-relevant information while suppressing irrelevant information (Chrysikou et al., 2014; Miller & Cohen, 2001; Shimamura, 2000). Performance should be optimal when the level of filtering is suitable to the demands of the task (Chrysikou et al., 2014). Consequently, increased top-down control can incur both costs and benefits. This is the case in instructed RL, where instruction-induced bias has been shown to vary according to individual differences in PFC dopaminergic tone caused by the catechol-*O*-methyltransferase (COMT) Val158Met genetic polymorphism (Doll et al., 2011). In particular, the Met allele, which has been shown to confer benefits in tests of working memory and cognitive control as compared to the Val allele (Durstewitz & Seamans, 2008; Witte & Flöel, 2012), is associated with a cost in the form of increased adherence to inaccurate instructions.

The goal of the present study was threefold. First, we sought to replicate the effect of COMT on instructed reinforcement learning, providing further evidence for the role of PFC-mediated top-down control in biasing RL.

Second, we aimed to expand the understanding of the impact of striatal dopaminergic genes on instructed RL. While Doll et al. (2011) examined the effects of genetic

polymorphisms specific to approach or avoidance learning in the striatum, we examined the effect of the DAT1/SLC6A3 variable number tandem repeat (VNTR), which affects striatal dopamine (DA) reuptake by altering expression of the dopamine transporter (DAT; Faraone et al., 2014; Vandenberg et al., 1992). Though there are conflicting reports on the exact effects of the DAT1/SLC6A3 VNTR, a recent meta-analysis suggests that in healthy individuals the 9-repeat allele is associated with increased DAT expression in human striatum, and thus potentially more efficient reuptake of DA as compared to the 10-repeat variant (Faraone et al., 2014; cf. Costa et al., 2011). Striatal DA levels have previously been linked to cognitive flexibility (Beeler, Daw, Frazier, & Zhuang, 2010; Cools & D'Esposito, 2009; Garcia-Garcia, Barceló, Clemente, & Escera, 2010), making DAT1 a plausible modulator of instructed RL.

Finally, while genetic and neuroimaging evidence is compelling, it falls short of establishing a causal role for PFC in biasing RL. We therefore hoped to establish this causal link by directly modulating PFC via transcranial direct current stimulation (tDCS). In keeping with a costs/benefits framework, we predicted that anodal stimulation—which has been successfully applied to PFC in order to improve cognitive control (Cattaneo, Pisoni, & Papagno, 2011; Fregni et al., 2005; Karuza et al., 2016; Nozari & Thompson-Schill, 2013; Zaehle, Sandmann, Thorne, Jäncke, & Herrmann, 2011)—would lead to increased bias due to increased top-down regulation. Cathodal stimulation over PFC has produced inconsistent results in cognitive domains (Jacobson et al., 2012; Nozari, Woodard, & Thompson-Schill, 2014). However, supporting the costs/benefits framework, it has been linked to decreased working memory (Zaehle et al., 2011) and

selective attention (Nozari et al., 2014; Zmigrod, Zmigrod, & Hommel, 2016), but improved dual task performance (Filmer, Mattingley, & Dux, 2013) and cognitive flexibility (Chrysikou et al., 2013). Therefore we tentatively predicted that cathodal stimulation would lead to decreased bias due to decreased top-down control of RL.

2. Methods

2.1. Subjects

One-hundred twenty-six right-handed subjects (42 per condition, 80 female, $M_{\text{age}} = 22.20$ years) participated in the study, receiving \$20 in compensation, regardless of performance. Informed consent was obtained from each subject in accordance with the University of Pennsylvania IRB. Subjects were randomly assigned to stimulation condition. We excluded a total of 23 subjects from the analyses for failure to meet the performance criteria described in section 2.3 (9 anodal, 6 cathodal, 8 sham), for a final sample of 103 (65 female, $M_{\text{age}} = 21.84$ years). Of these subjects, genotyping failed for one subject. For the Val158Met single-nucleotide polymorphism (SNP) of the COMT gene (rs4680), frequencies per allele in the final sample were 34:53:15 (Val/Val:Val/Met:Met/Met). For the DAT1/SLC6A3 VNTR in the 3' untranslated region, frequencies per allele were 65:26:6:2:1:1:1 (10/10:9/10:9/9:10/11:8/9:8/8:6/10). Subjects were placed in a 10/10 group if they had two repeats of 10+; otherwise they were placed in a 9-repeat carrier group (67 10/10, 35 9c). Neither gene differed from Hardy-Weinberg equilibrium either across the whole sample (all $ps > .14$) or within racial/ethnic subgroups (all $ps > .15$; see Supplementary Tables 3–6 in Appendix A for sample demographic breakdown). There was no association between COMT and DAT genotypes ($p > .35$,

Fisher's Exact Test), nor were there any associations between the two genes and stimulation condition (all $ps > .3$). For the dopamine genotype composite, the distribution of subjects was: 25:43:27:7 (0:1:2:3). The composite was also not significantly associated with stimulation condition ($p = .09$).

2.2. Materials and procedure

カ A (0.8)	ポ B (0.2)
ゴ C (0.6)	セ D (0.4)
ネ E (0.6)	バ F (0.4)

Table 2.1. Stimuli (reward probabilities) for the instructed probabilistic selection task. Subjects are instructed that D is the best symbol.

Subjects completed an instructed probabilistic selection task (iPST), presented on a 13" laptop computer via PsychoPy (Peirce, 2009). This task required subjects to learn the value of symbols initially presented in 3 pairs (AB, CD, EF; Table 2.1). Within each pair, one symbol had a higher probability of reward. Symbols were rendered as Japanese Hiragana characters, and the assignment of Japanese character to underlying stimulus was randomized across subjects. During the instructions, each symbol was presented individually for 5 seconds to familiarize subjects with the stimuli. Crucially, when

symbol D was presented the screen also displayed the following false advice: “This symbol has the best chance of being correct.”

During the training phase, subjects had to learn the value of each symbol via probabilistic feedback, which was delivered according to the symbol’s underlying $P(\text{reward})$.

Importantly, subjects were expected to learn to select the more highly rewarded symbol within each pair. Subjects completed 4 training blocks. Each block contained 20 repetitions of each pair, for a total of 60 trials per block and 240 total training trials. Trial order and feedback were randomized within each block. During the test phase, all possible symbol pairings were presented (e.g., AB, AC, AD, AE, AF, ...) without feedback. Each pair was presented 6 times, for a total of 90 trials. Order was randomized across subjects. See section 1.1 of Appendix A for further details regarding task design and presentation.

2.3. Performance criteria

Subjects had to meet the following performance criteria for the uninstructed symbols in order to be included in the analyses: $\geq 60\%$ accuracy on the AB pair and $\geq 50\%$ accuracy on the EF pair in at least one training block after the first block, with both criteria met in the same block. These criteria are similar to training phase learning criteria used in previous reports (Doll et al., 2011, 2009, 2014; Frank, Moustafa, Haughey, Curran, & Hutchison, 2007), but were relaxed slightly for AB to allow for additional variability in learning performance, given a previous report of tDCS effects on this pair (Turi et al., 2015). Subjects were also excluded if they failed to respond on $> 10\%$ of training trials.

In addition to excluding subjects who failed to pay attention or learn, these criteria helped ensure that subjects with arbitrary biases for one of the uninstructed symbols were excluded from the analyses. However, to further protect against arbitrary affinities introducing bias into the between-group analyses, we further tested for the presence of genotype or stimulation differences in the first 10 training trials of the uninstructed training pairs. There were no significant effects (all $ps > .10$), indicating that none of our genotype or stimulation groups entered the training phase with arbitrary stimulus preferences.

2.4. Genotyping

DNA samples were collected via Oragene saliva kits (DNA Genotek) and genotyped at the Penn Molecular Profiling Facility using standard procedures (see section 1.2 of Appendix A).

2.5. Transcranial direct current stimulation

We delivered direct current via saline-soaked sponge electrodes with a 25 cm² surface area. Current was generated by a continuous current stimulator (Magstim Eldith 1 Channel DC Stimulator Plus, Magstim Company Ltd., Whitland, Wales). In all conditions, 1.0 mA direct current was applied after a 30 second ramp-up period and was followed by a 30 second ramp-down. In the verum conditions, current was applied for 20 minutes. Stimulation was applied for only 30 seconds during sham. In the anodal condition, the anode was placed over F7, in accordance with the 10–20 international

system, and the cathode was placed over the right supraorbital. This placement was reversed in the cathodal condition.

Phase	tDCS	Duration
Instructions	No	Variable
Fixation	Yes	3 min
Training	Yes	17 min
Test Instructions	No	Variable
Test	No	6 min

Table 2.2. Stimulation procedure and duration for verum stimulation (sham was identical except stimulation only lasted for 30 seconds, at the onset of the fixation period).

The F7-RSO montage was chosen because current modeling (HDExplore Software, v2.3, SOTERIX) suggested it would maximize current through DLPFC sites found to be active during instructed reinforcement learning conditions (Fouragnan et al., 2013; Li et al., 2011). Stimulation at F7 has been shown to modulate prefrontally-mediated cognitive control across a range of tasks (Chrysikou et al., 2013; Lupyan, Mirman, Hamilton, & Thompson-Schill, 2012; Nozari et al., 2014). The procedure for each subject is outlined in Table 2.2. Stimulation began 180 seconds prior to the start of the first trial while subjects were presented with a fixation cross. Stimulation has not been shown to produce after-effects at 1.0 mA unless applied for at least 3 minutes, and thus this period gives stimulation time to take full effect (Nitsche & Paulus, 2000). Additionally, though stimulation ended after the training phase, after-effects have been reported up to an hour after stimulation lasting 9–13 minutes, so it is possible tDCS could directly affect

performance at test in addition to its indirect effect through modifying performance during training (Nitsche et al., 2003; Nitsche & Paulus, 2001).

2.6. Data analysis

Statistical analyses were conducted in R (R Core Team, 2018) using logistic mixed models implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015b). By modeling both fixed and random effects, these models controlled for the nonindependence inherent in within-subjects data. All models included random intercepts for subjects and random slopes for within-subjects variables and their interactions (Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009). When making between group comparisons of factors with more than two levels without planned comparisons, the significance of main effects and interactions were computed using the car package (Fox & Weisberg, 2011). Post-hoc comparisons were computed using the lsmeans package (Lenth, 2016). Significance levels for post-hoc comparisons were corrected using the Bonferroni-Holm method (Holm, 1979). Permutation tests were conducted via Monte Carlo sampling ($1.0e6 - 1$ permutations) using the perm package (Fay & Shaw, 2010).

2.7. Computational modeling

Reinforcement learning models were fit to each subject's data in order to evaluate hypotheses regarding the mechanisms of instructional bias. Models were fit by maximizing the log likelihood of the data using MATLAB's fmincon (Mathworks, MA, USA). To avoid local minima, each model fit was repeated 25 times from different

random starting points, using RMSEARCH. All models were fit to both training and test phase data. For the training phase, fits were optimized to account for subjects' trial-wise choices; for the test phase, they were optimized to result in learned Q-values after training that best account for choices during test (Frank et al., 2007).

Standard model. This model implements a standard Q-learning model with separate learning rates for gains and losses (Frank et al., 2007). The value of each stimulus is updated according to the following learning rule:

$$Q_{t+1}(s) = Q_t(s) + [\alpha_g \times \delta_t]_+ + [\alpha_l \times \delta_t]_-.$$

$$\delta_t = r_t - Q_t(s)$$

where $Q_t(s)$ is the action value of stimulus s at trial t , r_t is the reward (0 for losses, 1 for gains), and δ_t is the reward prediction error. The learning rate α_g applies only to gain trials, while the learning rate α_l applies only to loss trials.

Choice in the standard model and subsequently described models was implemented via a softmax function:

$$P_t(s_1) = \frac{\exp\left(\frac{Q_t(s_1)}{\beta}\right)}{\exp\left(\frac{Q_t(s_1)}{\beta}\right) + \exp\left(\frac{Q_t(s_2)}{\beta}\right)}$$

where $P_t(s_1)$ is the probability of choosing symbol s_1 over symbol s_2 , and β is a temperature parameter determining the extent to which choice is deterministic versus random.

For this model and subsequent models, we placed the following bounds on the parameters: $\alpha \in [0.002, 1]$; $\beta \in [0.06, 20]$. The temperature parameter was additionally

constrained by an empirical prior (Gershman, 2016): $\frac{1}{\beta} \sim \text{Gamma}(5.09, 0.83)$. Q-values for all stimuli were initialized at 0.5.

Learning bias model (Doll et al., 2011). The learning bias model is identical to the standard model in all respects except that when symbol D is chosen, the baseline learning rate is distorted as follows:

$$Q_{t+1}(D) = Q_t(D) + [\alpha_g \times \alpha_{bg} \times \delta_t]_+ + \left[\frac{\alpha_l}{\alpha_{bl}} \times \delta_t \right]_-$$

where α_{bg} increases the learning rate for instruction-consistent feedback (gains), α_{bl} diminishes the learning rate for instruction-inconsistent feedback (losses), and $\alpha_b \in [1, 10]^1$.

Bayesian hypothesis testing model (Doll et al., 2011). This model accounts for the possibility that the bias lies not in learning the value of the instructed stimulus D but in the decision to choose D. In this case, the choice bias requires that learners achieve a certain level of confidence that D is rewarded below chance before they abandon it. This model implements a Bayesian Q-learner with $Q_t(s) \sim \text{Beta}[\alpha_t(s), \beta_t(s)]$. After reward feedback, posterior Q-value distributions are updated as:

$$Q_{t+1}(s) \sim \text{Beta}[\alpha_t(s) + r_t, \beta_t(s) + (1-r_t)]$$

which increments α by 1 after gains and β by 1 after losses. Additionally, after every trial the α and β counts decay toward uniform, controlled by free parameters γ_α and γ_β ; $\gamma \in$

¹ In order to prevent learning rates exceeding 1.0, the learning bias parameters were also constrained such that $\alpha_b \leq \alpha^{-1}$ (Doll et al., 2009).

$[0, 1]$ (0 is full decay and 1 is no decay). Choice is implemented by submitting the mean of each symbol's beta distribution to the softmax function above. Crucially, when the instructed stimulus is encountered, a decision bias is implemented as follows:

$$P_t(s_{\text{alt}}) = \frac{\exp\left(\frac{0.5}{\beta}\right)}{\exp\left(\frac{0.5}{\beta}\right) + \exp\left(\frac{\mu_t(D) + \phi\sigma_t(D)}{\beta}\right)}$$

with $\phi \in [0, 20]$ and $P_t(D) = 1 - P_t(s_{\text{alt}})$. This decision rule dictates that the mean value of D must be greater than ϕ standard deviations of D below chance before it is more probable that the alternative symbol, s_{alt} , is chosen. Thus the more certain the learner is of the value of D , the lower the bias.

Decision bias model. Though the Bayesian hypothesis testing model has provided a reasonable fit to some subjects' training data and has been shown to be sensitive to individual differences, it has not overall outperformed the standard model in explaining training phase performance (Doll et al., 2011, 2009). It also compares the value of D to chance instead of to the value of the alternative stimulus, making it less effective as a possible model of test phase performance. Furthermore, interpretation of this model in comparison to the standard uninstructed model is complicated by the fact they are not nested models. Therefore, we also implemented a novel decision bias model. This model uses the standard Q-learner described above, but the softmax decision rule is modified for choices involving the instructed stimulus in a manner similar to the hypothesis testing model:

$$P_t(s_{\text{alt}}) = \frac{\exp\left(\frac{Q_t(s_{\text{alt}})}{\beta}\right)}{\exp\left(\frac{Q_t(s_{\text{alt}})}{\beta}\right) + \exp\left(\frac{Q_t(D) + \rho}{\beta}\right)}$$

with $\rho \in [0, 1]$. The free parameter ρ determines how much greater the value of the alternative symbol must be before it is more probable that it is chosen over D. Therefore, unlike the Bayesian model, this model: a) assumes a fixed bias; b) compares the value of D to the alternative symbol, making it more appropriate as a model of test phase choice; and c) contains the standard model as a special case ($\rho = 0$), ensuring differences in fit will be attributable to the presence of the bias and not to differences in the learner.

2.7.1. Model comparison

Goodness of fit was assessed using Akaike information criteria (AIC). We additionally submitted the AIC values to a Bayesian random effects analysis, which assumes there is a distribution of models in the population and attempts to identify which model is most prevalent. The quantity resulting from this analysis is a protected exceedance probability (PEP), which is the probability that a given model is the most frequent in the population, above and beyond chance (Rigoux, Stephan, Friston, & Daunizeau, 2014). PEPs were computed using the VBA toolbox (Daunizeau, Adam, & Rigoux, 2014). Model comparison was then made on the basis of both AIC and PEPs.

3. Results

We begin by reviewing general performance across the sample. We then examine genotypic differences in instructional bias. To this end, we first attempt to replicate the effect of COMT genotype. We then extend the investigation of the influence of dopaminergic genes on instructional bias to the DAT1 gene. In brief, we partially replicated the effect of COMT and found effects of DAT1 on instructional bias as well.

Motivated by these findings, we next ask whether a dopamine composite variable constructed from the COMT and DAT variables captures additional aspects of performance. These analyses demonstrated an overall graded effect of the dopamine composite on performance, and also uncovered a small group of subjects who demonstrated more extreme bias. We then ask if we can causally manipulate instructional bias with tDCS, finding that anodal stimulation had a small but significant effect on performance during training. Finally, we fit computational models to test potential mechanisms underlying instructional bias, finding evidence in favor of a model incorporating a bias on the decision to choose the instructed stimulus, rather than a bias on the learned value of the instructed stimulus.

3.1. General performance: Training phase

3.1.1. Instructed learning

We first conducted analyses of choice behavior during training. In all analyses, accuracy was binary coded (incorrect: 0, correct: 1), where correct is defined as choosing the stimulus with the higher probability of reward, regardless of whether it was rewarded on that trial. Trial Type was treatment coded (CD: 0, EF: 1). This coding allows direct assessment of how much instruction biased learning. Block was reverse Helmert coded in order to capture learning-related changes in the mean level of responding across training (i.e., Block 2 was compared to Block 1, Block 3 was compared to the mean of Blocks 1 and 2, and Block 4 was compared to the mean of all prior blocks). We assessed the effects of genotype and stimulation both by examining performance on the CD trials

alone, and by contrasting performance on CD with the equally rewarded but uninstructed EF pair. Given our between-subjects design, this latter contrast serves to account for additional variance in learning unrelated to instructional control. Therefore, instructed training models included all two-way and three-way interactions of genotype or tDCS condition, Trial Type, and Block.

Subjects were below chance on the CD pair ($\beta = -0.27$, $z = -2.86$, $p = .004$). Performance was significantly better on the EF pair ($\beta = 0.93$, $z = 6.96$, $p < .0001$), validating the success of the instructional manipulation. Despite poor overall performance on the CD pair, subjects continued to learn away from the instructions throughout training, as demonstrated by the significance of all three Block regressors (Block 2 vs. 1: $\beta = 0.32$, $z = 2.81$, $p = .005$; Block 3 vs. (1,2): $\beta = 0.24$, $z = 2.34$, $p = .02$; Block 4 vs. (1,2,3): $\beta = 0.31$, $z = 3.13$, $p = .002$).

3.1.2. Uninstructed learning

Variable coding in uninstructed training models was the same as above, except Trial Type was effect coded (AB: 1, EF: -1). The three-way interactions were not included in these models as there were no hypotheses relevant to these contrasts.

Subjects performed significantly above chance on uninstructed trials ($\beta = 1.12$, $z = 14.67$, $p < .0001$). There was an effect of Trial Type ($\beta = 0.46$, $z = 10.55$, $p < .0001$), indicating that subjects performed significantly better on the AB pair over the EF pair, in line with the relative difficulty of the two discriminations. Subjects continued to learn throughout training, though the magnitude of this effect was numerically smaller in later blocks

(Block 2 vs. 1: $\beta = 0.40$, $z = 4.79$, $p < .0001$; Block 3 vs. (1,2): $\beta = 0.22$, $z = 2.74$, $p = .006$; Block 4 vs. (1,2,3): $\beta = 0.18$, $z = 2.17$, $p = .03$). There was additionally a Trial Type x Block 2 vs. 1 interaction ($\beta = 0.31$, $z = 3.97$, $p < .0001$), indicating a steeper learning trajectory for AB over EF during the initial blocks of the task, which again is unsurprising given the relative ease of the AB discrimination.

3.2. General performance: Test phase

The training and test phases are purported to represent different processes subserved by different neural systems (Frank et al., 2007). While the training phase is supposed to reflect hippocampally- and frontally-mediated memory and hypothesis-testing processes, the test phase is designed to give a “purer” measure of striatally-learned reinforcement values. The standard approach to assessing performance at test is to examine performance on trials in which a stimulus of interest is included in novel pairings, giving an estimate of how well underlying reward values were learned during training.

Two measures from the literature were used to assess the effect of instruction on test phase performance (Doll et al., 2014). The first analysis compared performance on Avoid-D (AD, DE) vs. performance on Avoid-F (AF, CF). For both measures, the target stimulus should not be chosen, as it has been paired with stimuli that had a higher probability of reward during training. Given that D and F had identical reward probabilities during training, subjects should perform equally well on both measures. However, if instruction biased the ultimate reward values subjects learned, or if subjects’

choices continue to be biased at test, they should avoid D at a lower rate than they avoid F.

Choice Type was entered as an effect-coded factor (Avoid-D: 1, Avoid-F: -1) in a logistic mixed effects model of choice performance. The intercept was significant ($\beta = 0.73, z = 5.53, p < .0001$), indicating that subjects' overall avoidance of D and F was above chance. There was also a main effect of Choice Type ($\beta = -0.61, z = -5.43, p < .0001$). As expected, subjects showed a confirmation bias effect, avoiding D significantly less than they avoided the equally rewarded symbol F.

The second analysis of instructed learning examined performance on DF trials in order to directly compare the relative subjective value of the two stimuli. A greater effect of instruction on learning, and thus a greater bias, should be associated with an increased tendency to choose D over F.

In this model, choice on DF trials was the dependent variable (D: 1, F: 0). The intercept was significant ($\beta = 1.58, z = 5.86, p < .0001$). Subjects demonstrated a strong bias—they were almost five times more likely to choose D, as indicated by an odds ratio (OR) of 4.86. In sum, our training and test results replicate previous investigations (Doll et al., 2011, 2009, 2014) and confirm that the instructional manipulation was successful.

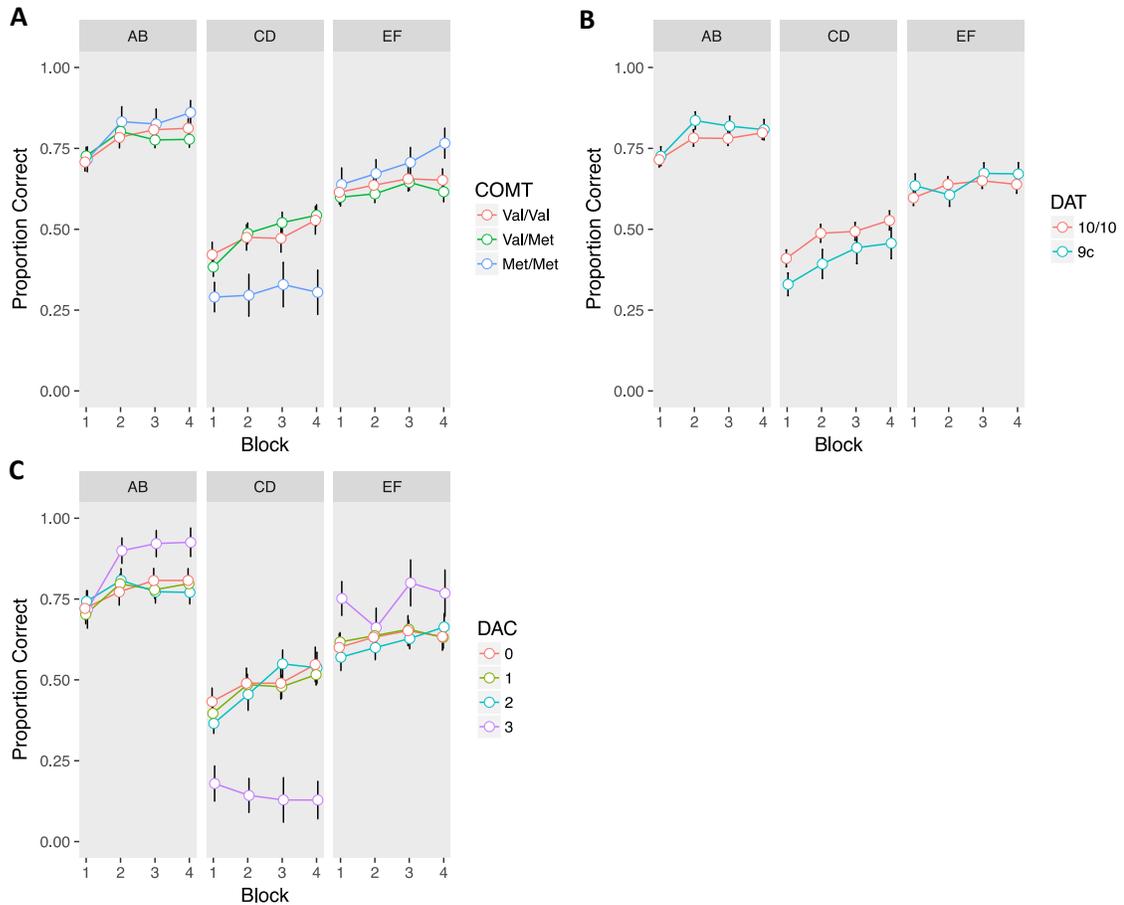


Figure 2.1. Training phase performance by trial type (AB, CD, EF) and genotype. Accuracy is defined as the proportion of time the symbol with the higher reward probability was chosen, regardless of whether it was rewarded or not. Error bars are standard errors of the mean. **A** COMT, **B** DAT, **C** Dopamine composite (DAC).

3.3. COMT: Training phase

3.3.1. Instructed learning

We next sought to replicate the effect of the COMT Met allele on adherence to the instructions during training (Doll et al., 2011). COMT genotype was effect coded. All other variables were coded as above.

There was a significant COMT x Trial Type interaction ($\chi^2(2) = 13.94, p = .0009$). Met homozygotes were significantly worse overall on the instructed pair (Figure 2.1A, Supplementary Table 7), as compared to both heterozygotes ($\beta = -0.98, z = -3.70, p_{corrected} = .001$) and Val homozygotes ($\beta = -0.92, z = -3.26, p_{corrected} = .006$). Met homozygotes also demonstrated better performance at a trend level on the uninstructed EF pair compared to Val/Met subjects, but this did not survive correction for multiple comparisons ($\beta = 0.41, z = 1.87, p = .06, p_{corrected} = .25$). Notably, no other comparisons reached significance, including the comparison of instructed performance between Val/Met and Val/Val subjects (all $ps > .2$), indicating impaired performance was specific to Met homozygotes.

Because our Met/Met group was somewhat small ($N = 15$) due to the low frequency of this genotype in the general population (Auton et al., 2015), we took a number of additional steps to ensure these results were not spurious. First, we reran our analyses comparing Val homozygotes to Met carriers (Metc), which was also the analysis performed by Doll and colleagues (2011). In this case, we failed to replicate the effect of Met-carrier status on instructed learning. The Metc x Trial Type interaction was not significant ($\chi^2(1) = 0.16, p = .69$), nor were there any other significant effects of Met carrier status (all $ps > .42$). We then asked whether the full COMT model or the Metc model provided a better fit to the data, finding that the COMT model was a modestly better fit, despite including additional parameters ($AIC_{COMT} = 19966, AIC_{Metc} = 19969$). Finally, we conducted permutation tests on CD trials, averaged across all blocks, to further guard against the possibility that our Met homozygote results could have arisen

under the null. Confirming our results, Met homozygotes' performance was reliably below the mean on CD trials ($p = .004$), and this group performed worse than both Val homozygotes ($p_{corrected} = .006$) and heterozygotes ($p_{corrected} = .001$). We therefore utilize the full breakdown of COMT genotype for the remainder of the results.

3.3.2. Uninstructed learning

In contrast to instructed learning, we found no effects of COMT genotype on uninstructed learning (all $ps > .2$; Figure 2.1A, Supplementary Table 8).

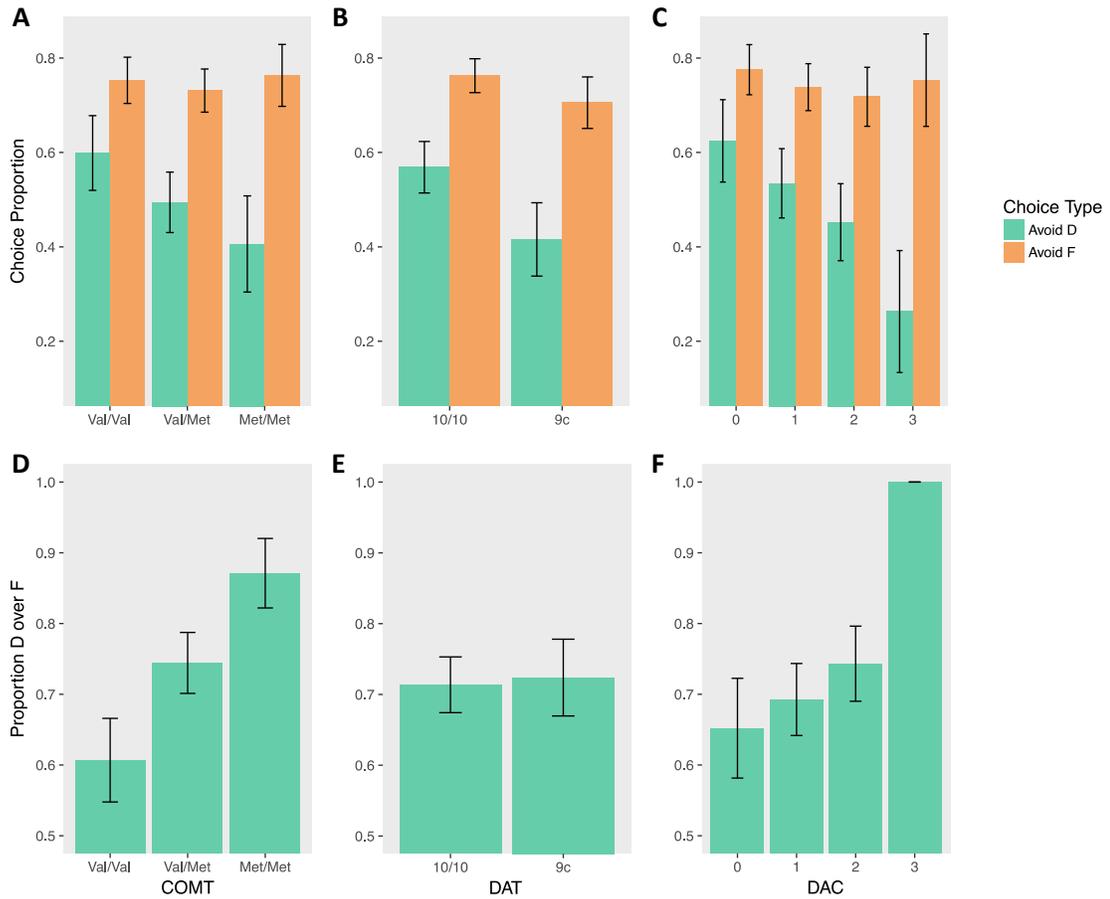


Figure 2.2. Test phase performance by genotype. **Top** Accuracy avoiding D (instructed) and F (uninstructed) when paired with stimuli at test that had a higher reward probability

during training for **A** COMT, **B** DAT, **C** DAC. **Bottom** Proportion by which D was chosen over F at test for **D** COMT, **E** DAT, **F** DAC.

3.4. COMT: Test phase

Instructed test phase performance demonstrated evidence of a gene-dose effect (Figure 2.2A,D). COMT status significantly predicted performance on DF trials ($\chi^2(2) = 9.06, p = .01$). Val homozygotes were less likely to choose D on DF trials compared to heterozygotes ($\beta = -1.11, z = -2.09, p_{corrected} = .07$) and to Met homozygotes ($\beta = -2.28, z = -2.82, p_{corrected} = .01$). There was no significant difference between Val/Met and Met/Met groups, but Met/Met subjects were numerically more likely to choose D ($\beta = 1.18, z = 1.53, p_{corrected} = .13$). Supporting this pattern, an exploratory gene-dose analysis demonstrated a significant linear effect of the number of Met alleles on choosing D over F ($\beta = 1.60, z = 3.01, p = .003$).

There were no significant effects of COMT genotype on the Avoid-D/Avoid-F measure, (all $ps > .17$), but quantitatively, differences were indicative of a similar gene-dose relationship on Avoid-D. An exploratory gene-dose analysis demonstrated a trend-level Met x Trial Type interaction ($\beta = -0.44, z = -1.86, p = .06$). While increasing Met alleles negatively predicted performance on Avoid-D ($\beta = -0.90, z = -2.09, p_{corrected} = .07$), there was no relationship with uninstructed Avoid-F ($\beta = -0.02, z = -0.05, p_{corrected} = .96$).

The above results refine, but only partially replicate, the effect of COMT genotype on instructed RL. While Doll and colleagues (2011) found that Met carriers demonstrate greater instructional bias relative to Val homozygotes during training, we found increased

bias exclusively for Met homozygotes. Our COMT test phase results provide novel evidence for a gene-dose effect, though differences on the Avoid-instructed measure were not as robust as reported previously. The prior report included a somewhat greater percentage of Met homozygotes out of all Met carriers (28.3%) than the present study (22.1%), which could have impacted the results given that instructional bias appears to be strongest in the former group. Additionally, a number of methodological differences could have contributed to these discrepancies. These differences aside, as COMT is thought to be particularly and differentially important to the regulation of prefrontal dopamine levels (Durstewitz & Seamans, 2008; Tunbridge, 2010), the present findings further implicate prefrontal cortex in biasing responding to instructed stimuli at both training and test.

3.5. DAT: Training phase

Expanding the investigation of the effect of dopaminergic genes on instructional bias, we next examined the effect of DAT1 genotype. In our regression models, DAT was simple coded with 10/10 homozygotes as the reference (9c: 0.5, 10/10: -0.5).

As compared to 10-repeat homozygotes, 9-repeat carriers were significantly worse on the instructed pair ($\beta = -0.43$, $z = -2.17$, $p = .03$; Figure 2.1B, Supplementary Table 9).

There was also a trend-level DAT x Trial Type interaction ($\beta = 0.53$, $z = 1.91$, $p = .056$).

While 9-repeat carriers were worse on the CD pair, there was no difference between genotypes on the EF pair ($p > .5$). There were no interactions between DAT and Block, indicating similar learning trajectories in both groups (all $ps > .4$). Nor were there any

effects of DAT on uninstructed learning (all p s > .4; Figure 2.1B, Supplementary Table 10).

3.6. DAT: Test phase

There was no effect of DAT on DF trials ($p = .74$; Figure 2.2E). There was a main effect of DAT on Avoid-D/Avoid-F ($\beta = -0.66$, $z = -2.40$, $p = .02$; Figure 2.2B) in the absence of a significant interaction ($p > .17$), suggesting that 9-repeat carriers were significantly worse overall on these measures. However, the effect seems to be driven primarily by worse performance on Avoid-D (Avoid-D: $\beta = -0.49$, $z = -2.30$, $p_{corrected} = .04$; Avoid-F: $\beta = -0.17$, $z = -1.18$, $p = .24$).

Remarkably, though DAT plays little role in cortical DA clearance (Sulzer, Cragg, & Rice, 2016), it appears to be equally if not more predictive of training than test phase performance, the former of which is putatively more reliant on prefrontal function (Frank et al., 2007). This result is surprising, given that investigations assessing other striatal genes have found that striatal genotypic effects in both instructed and uninstructed learning are confined to the test phase only (Doll et al., 2011; Frank et al., 2007).

Previous work has indicated that there is a reciprocal relationship between prefrontal and striatal DA, with more prefrontal DA leading to more cognitive stability, while more striatal DA leads to more cognitive flexibility (Cools & D'Esposito, 2009). Motivated by this and by prior studies in which composites of multiple DA genes have shown better predictive power than single genes (Kohno et al., 2016; Nikolova, Ferrell, Manuck, &

Hariri, 2011), we next asked whether a composite DA variable would better predict instructional bias.

3.7. DA composite: Training phase

To produce the DA composite (DAC), we recoded the COMT and DAT variables according to putative prefrontal-striatal DA balance (COMT: Val/Val = 0, Val/Met = 1, Met/Met = 2; DAT: 10/10 = 0, 9c = 1), and then summed the two variables. The resulting composite ranged between 0 (low frontal DA, high striatal DA) and 3 (high frontal DA, low striatal DA).

Reexamining training phase performance (Figure 2.1C, Supplementary Table 11), we found a significant effect of DAC ($\chi^2(3) = 11.02, p = .01$), superseded by a significant DAC x Trial Type interaction ($\chi^2(3) = 29.56, p < .0001$). Post-hoc comparisons revealed that the DAC3 group was significantly and uniquely impaired in learning away from the instructions compared to the other three groups (DAC3 vs. DAC 0: $\beta = -2.03, z = -5.52, p_{corrected} < .0001$; DAC3 vs. DAC 1: $\beta = -1.96, z = -5.56, p_{corrected} < .0001$; DAC3 vs. DAC 2: $\beta = -1.99, z = -5.46, p_{corrected} < .0001$). In contrast, DAC3 subjects demonstrated better performance on EF, though this did not survive correction for multiple comparisons: (DAC3 vs. DAC 0: $\beta = 0.62, z = 1.94, p = .053, p_{corrected} = .42$; DAC3 vs. DAC 1: $\beta = 0.58, z = 1.92, p = .056, p_{corrected} = .42$; DAC3 vs. DAC 2: $\beta = 0.68, z = 2.14, p = .03, p_{corrected} = .29$). No other comparisons between DAC groups were significant (all $ps > .6$). The DAC x Trial Type interaction was already present in the first block of training, suggesting it was not the result of extensive learning ($\chi^2(3) = 15.89, p = .001$).

Nor was it ameliorated by additional training, as the DAC3 group was the only group to show no evidence of learning away on CD from the first block to the last (DAC 0: $\beta = 0.53$, $z = 2.00$, $p_{corrected} = .09$; DAC 1: $\beta = 0.57$, $z = 2.83$, $p_{corrected} = .01$; DAC 2: $\beta = 0.80$, $z = 3.11$, $p_{corrected} = .008$; DAC 3: $\beta = -0.59$, $z = -1.03$, $p_{corrected} = .30$).

There were no significant differences between DAC groups in the analysis of uninstructed learning (all $ps > .3$), though as with EF, the DAC3 group's performance was quantitatively better on AB (Figure 2.1C, Supplementary Table 12). These differences in uninstructed learning are intriguing given that they are in the opposite direction of the instructed effect, but given the small sample size of the DAC3 group ($N = 7$) due to the lower prevalence of both the COMT Met allele (Auton et al., 2015) and the DAT 9-repeat variant (Doucette-Stamm, Blakely, Tian, Mockus, & Mao, 1995; Vandenberg et al., 1992) in the general population, this study may not have had the statistical power to determine whether such small effects are reliable.

As with the COMT Met/Met results, because of the small sample size of the DAC3 group, we again took efforts to ensure these results did not arise by chance. First, we repeated the analysis with a modified DA composite created by summing the Metc and DAT variables (Metc: Val/Val = 0, Met carrier = 1; DAT: 10/10 = 0, 9c = 1), producing three DACmetc groups $Ns = 25:51:26$ (0:1:2). Repeating our analysis of instructed learning, we failed to find any effects of DACmetc (all $ps > .21$). However, the full DAC model provided a much better fit to the data, despite including additional parameters ($AIC_{DAC} = 19958$, $AIC_{DACmetc} = 19974$), and also provided a better fit than both the COMT and DAT instructed learning models ($AIC_{COMT} = 19966$, $AIC_{DAT} = 19961$).

Permutation tests on the average performance on CD trials across training also support the results of the regression analysis. DAC3 subjects were reliably below the mean on CD trials ($p < .0001$), and this group performed worse than all other DAC groups (DAC3 vs. DAC0: $p_{corrected} = .0001$, DAC3 vs. DAC1: $p_{corrected} < .0001$, DAC3 vs. DAC2: $p_{corrected} = .0001$). Given that it is highly unlikely that seven randomly chosen subjects would have performance at the level of the DAC3 group, we utilize the full DAC composite for the remainder of the results.

3.8. DA composite: Test phase

While there was only a marginal main effect of DAC on Avoid-D/Avoid-F ($\chi^2(3) = 6.62$, $p = .085$), a gene-dose analysis revealed a significant linear effect of DAC ($\beta = -0.86$, $z = -2.56$, $p = .01$) qualified by a DAC x Choice Type interaction ($\beta = -0.62$, $z = -2.16$, $p = .03$). DAC status was negatively associated with avoiding D; it showed no relationship to avoiding F (Avoid-D: $\beta = -1.47$, $z = -2.88$, $p_{corrected} = .008$; Avoid-F: $\beta = -0.24$, $z = -0.68$, $p_{corrected} = .50$; Figure 2.2C). DF trials revealed a similar pattern; though there was no main effect of DAC ($\chi^2(3) = 1.12$, $p = .77$), there was a significant gene-dose effect, with increasing choice of the instructed stimulus with increasing DAC status ($\beta = 1.72$, $z = 2.53$, $p = .01$). This effect appears to be driven primarily by the DAC3 group, all seven of whom remarkably chose D over F 100% of the time (Figure 2.2F).

In sum, there was graded effect of DAC on test phase performance, with increasing frontal (decreasing striatal) DA predicting greater adherence to the instructions. This

graded relationship was punctuated by the performance of the DAC3 group, who, as during training, demonstrated substantially greater instructional bias.

Taken together, the genotyping results implicate prefrontal cortex, and in particular the balance between prefrontal and striatal dopamine, in modulating instructed RL. This pattern motivates asking our next question: Does experimentally manipulating prefrontal function via tDCS alter the magnitude of instructional bias?

3.9. tDCS: Training phase

3.9.1. Instructed learning

To examine the main hypotheses of the study—that anodal stimulation will increase confirmation bias, while cathodal stimulation may decrease it—our focal analyses concerned the contrasts of Anodal vs. Sham stimulation and Cathodal vs. Sham stimulation. These contrasts include Condition, or the overall effect of stimulation compared to Sham on instructed choice, and Condition x Trial Type, which allows for the same assessment while controlling for performance on EF. For a more fine-grained investigation of the time course of learning, we additionally examined the Condition x Block interactions, which indicate whether stimulation altered the extent to which subjects learned away from the instructions across training blocks, and the Condition x Trial Type x Block interactions, which allow for the same assessment while controlling for performance on EF. Condition was simple coded with sham as the reference (Anodal: $2/3 - 1/3$, Cathodal: $-1/3 2/3$, Sham: $-1/3 -1/3$).

We first examined the contrasts between anodal and sham stimulation. Supporting our hypothesis, there was a significant Anodal vs. Sham x Trial Type x Block 2 vs. 1 interaction ($\beta = 0.76, z = 2.22, p = .03$). When controlling for performance on EF, the sham group demonstrated significant learning away from the instructions from Block 1 to Block 2 on CD, while the anodal group did not (Sham: $\beta = 0.63, z = 2.60, p_{corrected} = .046$; Anodal: $\beta = -0.13, z = -0.53, p_{corrected} = 1.00$). The sham group nearly doubled their performance (OR = 1.88), but the anodal group demonstrated essentially no learning (OR = 0.88; Figure 2.3A,B and Table 2.3). Examining performance on CD without adjusting for EF, the Anodal vs. Sham x Block 2 vs. 1 interaction was at trend ($\beta = -0.45, z = -1.59, p = .11$). As above, the sham group showed significant learning from Block 1 to Block 2, while the anodal group did not (Sham: $\beta = 0.53, z = 2.68, p_{corrected} = .04$; Anodal: $\beta = 0.08, z = 0.42, p_{corrected} = 1.00$). In contrast, neither group demonstrated significant learning from Block 1 to Block 2 on EF (Sham: $\beta = -0.10, z = -0.64, p_{corrected} = 1.00$; Anodal: $\beta = 0.21, z = 1.33, p_{corrected} = .74$).

Predictor	β	OR ^a	z	p
Intercept	-0.27	0.76	-2.86	.004
Anodal vs. Sham	-0.01	0.99	-0.04	.97
Cathodal vs. Sham	0.09	1.10	0.40	.69
Trial Type	0.93	2.53	6.96	< .0001
Block 2 vs. 1	0.32	1.38	2.81	.005
Block 3 vs. (1,2)	0.24	1.27	2.35	.02
Block 4 vs. (1,2,3)	0.31	1.36	3.13	.002
Anodal vs. Sham x Trial Type	-0.05	0.95	-0.15	.88

Cathodal vs. Sham x Trial Type	-0.17	0.84	-0.54	.59
Anodal vs. Sham x Block 2 vs. 1	-0.45	0.64	-1.59	.11
Anodal vs. Sham x Block 3 vs. (1,2)	0.18	1.20	0.73	.47
Anodal vs. Sham x Block 4 vs. (1,2,3)	0.08	1.07	0.31	.75
Cathodal vs. Sham x Block 2 vs. 1	-0.19	0.83	-0.69	.49
Cathodal vs. Sham x Block 3 vs. (1,2)	-0.10	0.91	-0.41	.68
Cathodal vs. Sham x Block 4 vs. (1,2,3)	-0.17	0.85	-0.70	.48
Trial Type x Block 2 vs. 1	-0.22	0.80	-1.58	.11
Trial Type x Block 3 vs. (1,2)	-0.03	0.97	-0.22	.82
Trial Type x Block 4 vs. (1,2,3)	-0.16	0.85	-1.11	.27
Anodal vs. Sham x Trial Type x Block 2 vs. 1	0.76	2.15	2.22	.03
Anodal vs. Sham x Trial Type x Block 3 vs. (1,2)	-0.25	0.78	-0.75	.46
Anodal vs. Sham x Trial Type x Block 4 vs. (1,2,3)	0.06	1.06	0.17	.86
Cathodal vs. Sham x Trial Type x Block 2 vs. 1	0.47	1.60	1.41	.16
Cathodal vs. Sham x Trial Type x Block 3 vs. (1,2)	-0.12	0.89	-0.37	.71
Cathodal vs. Sham x Trial Type x Block 4 vs. (1,2,3)	0.50	1.65	1.46	.14

Note. Boldfaced text indicates $p < .05$. ^aOR: Odds Ratio

Table 2.3. Mixed effects logistic regression model of the effect of instruction (CD vs. EF) and tDCS on training phase performance.

We also sought to ensure that the effect of anodal stimulation early in learning was not driven by the presence of DAC3 subjects. Controlling for DAC, the Anodal vs. Sham x Trial Type x Block 2 vs. 1 interaction remained significant ($\beta = 0.86$, $z = 2.45$, $p = .01$) and the Anodal vs. Sham x Block 2 vs 1 interaction for CD remained at trend ($\beta = -0.49$, $z = -1.74$, $p = .08$), confirming that the effect was not driven by genotypic differences between stimulation conditions.

Taken together, these results indicate that anodal stimulation significantly impeded learning away from the instructions during the initial blocks. No other Anodal vs. Sham contrasts were significant (Table 2.3), including the overall effect of anodal stimulation ($p = .97$) and the Anodal vs. Sham x Trial Type interaction ($p = .88$), suggesting that anodal stimulation only weakly and transiently affected performance. In contrast to the anodal condition, there were no significant effects of cathodal stimulation (all $ps > .14$).

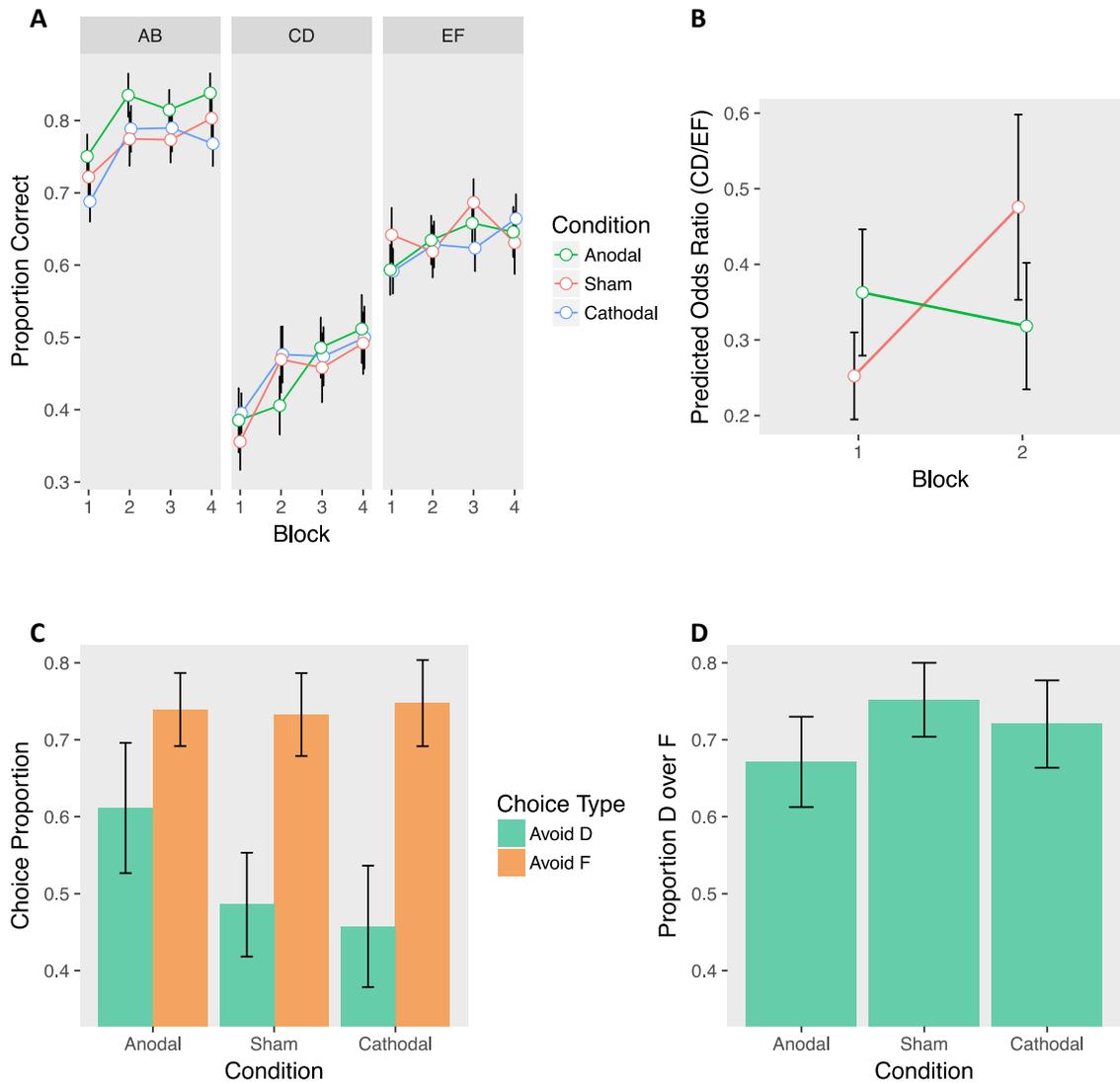


Figure 2.3. Performance at training (**top**) and test (**bottom**) by tDCS stimulation condition. **A** Training phase performance by trial type. **B** The effect anodal stimulation on instructed reinforcement learning. Points are predicted odds ratios for the CD/EF contrast by block and condition. This contrast reflects performance on CD controlling for performance on EF, giving a purer measure of the effect of instructions on choice. Lines represent the two-way Trial Type x Block interactions within each condition. Error bars are standard errors of the parameter estimates. While the sham group demonstrated significant learning away from the instructions from Block 1 to Block 2, the anodal group

did not, and this interaction was significant (see section 3.9.1). **C** Avoid-D/Avoid-F. **D** DF trials.

3.9.2. Uninstructed learning

We also explored the effect of stimulation on accuracy during training for the uninstructed symbol pairs (AB, EF). Quantifying the effect of stimulation on uninstructed learning is important in order to show that effects on instruction are not in some way due to generally altered learning, especially given a prior report of altered performance on the AB pair under anodal stimulation (Turi et al., 2015).

Though there were no significant effects of stimulation condition at the $p < .05$ level, there was a trend-level Anodal vs. Sham x Trial Type interaction ($\beta = 0.16$, $z = 1.66$, $p = .097$; Figure 2.3A, Supplementary Table 13), reflecting somewhat better average performance on the AB pair by the anodal group. This difference is intriguing given increasing evidence that working memory processes contribute to RL performance (Collins, Ciullo, Frank, & Badre, 2017; Collins & Frank, 2012), and anodal stimulation has been shown to improve working memory (Fregni et al., 2005; Nozari & Thompson-Schill, 2013; Zaehle et al., 2011). However, in light of the marginal nature of this unhypothesized effect, we do not interpret it further. As with instructed learning, there were no significant effects of cathodal stimulation (all $ps > .12$).

3.10. tDCS: Test phase

In contrast to the training phase, there were no significant effects of stimulation on either Avoid-D/Avoid-F or DF trials at test (all $ps > .19$; Figure 2.3C,D). This suggests that

unlike COMT genotype, to the extent that tDCS modulated instructed learning, it biased choice during training without impacting the learned value of the instructed stimulus.

3.11. Computational modeling

While the behavioral analyses above confirm the existence of instructional bias, they are only weakly informative with respect to the underlying mechanisms. Two classes of models have been suggested to account for instructional bias on the PST: models in which instructions bias striatal reward learning (learning bias models), and those in which instructions affect choice rather than learning (choice bias models) (Doll et al., 2009). Prior work has provided weak evidence for a choice bias operating during training, while test phase performance has been best explained by a learning bias mechanism (Doll et al., 2011, 2009). Two results from the present study bear on this question. First, the early-developing, persistent bias of the DAC3 group during training, coupled with their exclusive choice of D over F at test, would seem to be more consistent with a choice bias during both phases. However, these effects could also plausibly arise from a very strong learning bias, making this interpretation far from definitive. Second, the unaltered performance by the anodal group at test also appears more consistent with tDCS influencing a choice bias early in training, though caution is warranted in interpreting a null result.

We therefore fit computational models to subjects' data—one learning bias model and two choice bias models—each of which encapsulates a different hypothesis about the nature of instructional control (see section 2.7). Briefly, the *learning bias model* (Doll et

al., 2009) assumes instructional bias arises from an increase in learning rate for gains and a decrease in learning rate for losses when the instructed symbol D is selected. The *Bayesian hypothesis testing model* (Doll et al., 2009) assumes that subjects veridically learn the reward value of D in a Bayesian fashion, but must have a certain level of confidence that the value of D is below chance before they reliably stop choosing it. We additionally implemented a novel choice bias model, the *decision bias model*, which assumes a standard RL learner with a fixed bias added to the value of D during choice. Finally, we fit a standard RL model, which tests the null hypothesis of no bias.

Contrary to prior work, both the training and test phase were best explained by the decision bias model (Table 2.4). However, while AIC strongly supported this model at both training and test, the protected exceedance probabilities and estimated model frequencies did not provide strong evidence that this model was more frequent in the population for the training phase than the Bayesian hypothesis testing model. We therefore examined the correlation between each model's bias parameter and performance on CD trials across training, in order to ascertain whether one or the other model better accounted for behavior on instructed learning trials. The ϕ parameter of the Bayesian hypothesis testing model was significantly correlated with performance on CD trials ($r(101) = -.23, p = .02$). However, the correlation between the ρ parameter of the decision bias model and CD performance was much stronger ($r(101) = -.66, p < .0001$), and the difference between the correlations was significant (Steiger's $Z = -3.82, p = .0001$). In accordance with our tentative hypothesis based on the behavioral results, we

conclude that both training and test phase performance can be parsimoniously accounted for by a single choice bias mechanism.

Model	FP	-LL	AIC	PEP	EF
Training					
standard	3	13543.8	27705.7	0.007	0.239
learning bias	5	13298.2	27626.4	5.081e-05	0.003
Bayes HT	4	13402.9	27629.8	0.339	0.362
decision bias	4	13284.2	27392.4	0.654	0.397
Test					
standard	3	4755.4	10128.8	3.816e-16	0.164
learning bias	5	4362.8	9755.7	3.816e-16	0.002
Bayes HT	4	4748.4	10320.8	3.816e-16	0.015
decision bias	4	4304.6	9433.3	1.000	0.818

Table 2.4. Model comparison of reinforcement learning model fits to subject data. FP: number of free parameters; -LL: negative log-likelihood; AIC: Akaike information criteria; PEP: protected exceedance probability, the probability that a given model is the most frequent in the population, above and beyond chance; EF: estimated model frequency, the frequency of the model in the population as estimated by the Bayesian random-effects analysis.

We also reexamined genotypic and stimulation group differences with respect to the ρ parameter of the decision bias model. These results are reported fully in section 2 of Appendix A and average parameter estimates are reported in Supplementary Tables 1 and 2. Briefly, we found effects of COMT and DAC on ρ at both training and test, in the same direction as the behavioral results. For DAT, 9-repeat carriers were fit with a higher ρ parameter during training, but test phase differences were best explained by the 9-repeat carrier group being fit with a lower learning rate for losses as compared to 10/10 group. We were, however, unable to confirm the anodal tDCS behavioral effect in the

parameters of the decision bias model. While this does not invalidate the effect, it does warrant additional caution in interpreting the result.

4. Discussion

There is mounting evidence that reward learning is far more complex and dynamic than can be accounted for by simple model-free theories of reinforcement. This complexity has been explored with respect to goal-directed planning processes (i.e., model-based RL) (Dolan & Dayan, 2013) and instructional control (Wolfensteller & Ruge, 2012), among others. Both model-based RL and instructional control have been associated with cognitive control and frontostriatal function (Daw, Niv, & Dayan, 2005; Doll et al., 2011, 2009, 2014; Fouragnan et al., 2013; Li et al., 2011; Otto et al., 2015; Smittenaar et al., 2013; Wolfensteller & Ruge, 2012). While the importance of cognitive control to healthy cognitive functioning is indisputable, top-down control can be detrimental to learning and cognitive flexibility (Chrysikou et al., 2014; Gopnik et al., 2015).

In the case of instructed reinforcement learning, increased top-down control can be detrimental in that it leads to greater instructional bias toward inaccurate instructions. This study expands on the finding that instructional bias is associated with dopaminergic genes affecting PFC and striatal function (Doll et al., 2011), suggesting that the balance between PFC DA (COMT) and striatal DA (DAT1) modulates instructed learning. We further establish a causal link between PFC and biases found in instructed RL. In accord with our hypothesis, anodal subjects demonstrated more protracted learning away from the instructions during the early blocks of training, complementing the genetic evidence

that individual differences associated with PFC function are linked to individual differences in instructional control of RL.

4.1. A dopamine genetic composite is associated with instructed learning

While both COMT Met/Met genotype and DAT1 9-repeat carrier genotype were individually significant predictors of greater instructional control during training, the DA composite revealed that this effect was selective to Met/Met:9-repeat carriers (DAC3). This greater bias emerged early in training and persisted throughout the training phase, unaffected by feedback. During test, a gene-dose effect, confirmed both within each gene and with the composite, demonstrated greater bias with increasing Met alleles and decreasing DAT1 repeats. These results are consistent with the known reciprocal relationship between PFC and striatal DA (King, Zigmond, & Finlay, 1997; Kolachana, Saunders, & Weinberger, 1995; Meyer-Lindenberg et al., 2005). It has been hypothesized that the balance between cognitive stability and cognitive flexibility is mediated via corticostriatal interactions and the differential modulation of prefrontal and striatal circuits by DA. While increases in prefrontal relative to striatal DA have been linked to cognitive stability, increases in striatal relative to prefrontal DA have been linked to cognitive flexibility (Cools & D'Esposito, 2009). We propose that increasing PFC DA, indexed by increasing Met alleles, coupled with decreasing tonic striatal DA, indexed by decreasing DAT1 repeats, shifts the balance away from bottom-up striatal learning based on reward prediction errors and toward PFC-mediated top-down control of RL.

While extracellular DA is primarily recycled via reuptake by DAT in striatal regions, there is little DAT expression in PFC, where levels of DA are controlled by reuptake via the norepinephrine transporter (NET) and enzymatic breakdown via COMT (Seamans & Yang, 2004; Sulzer et al., 2016). With regard to COMT, PFC DA plays a critical role in stabilizing working memory representations (Durstewitz & Seamans, 2008), which are thought to facilitate top-down control (Miller & Cohen, 2001). Notably, carriers of the Met allele of the Val158Met genetic polymorphism have diminished COMT enzyme activity and concomitantly higher levels of prefrontal dopamine (see Tunbridge, 2010 for review). Elevated DA in PFC may then cause increased D1 receptor stimulation, which further drives activity in PFC afferents such as the striatum (Bilder, Volavka, Lachman, & Grace, 2004). Indeed, frontostriatal functional connectivity varies with COMT genotype (Krugel et al., 2009; Tan et al., 2007; Tunbridge, Farrell, Harrison, & Mackay, 2013). Behaviorally, the Met allele has been associated with enhanced working memory and cognitive control (see Witte & Flöel, 2012 for review). Carriers of the Val allele have more rapid breakdown of prefrontal dopamine and thus somewhat weaker working memory, but potentially greater cognitive flexibility (Krugel et al., 2009; Witte & Flöel, 2012). Replicating previous findings (Doll et al., 2011), the Met allele in our study was associated with greater instructional bias and therefore indicative of greater top-down control.

In the case of the DAT1/SLC6A3 VNTR, our behavioral results are consistent with increased DAT expression with the 9-repeat allele (Faraone et al., 2014) leading to reductions in tonic DA concentrations in the striatum. Reduced tonic DA in striatum has

been shown to facilitate PFC input (Goto & Grace, 2005), which would in turn allow for greater biasing of RL. Furthermore, human imaging studies have demonstrated that DAT1 and COMT affect activity in prefrontal and striatal regions during reward anticipation. While the results of these studies are not entirely consistent, anticipatory activity in striatum is generally greater for DAT1 9-repeat carriers and is modulated by COMT genotype (Aarts et al., 2010; Dreher, Kohn, Kolachana, Weinberger, & Berman, 2009; cf. Yacubian et al., 2007), with one study finding the highest activity in both lateral PFC and ventral striatum for Met/Met:9-repeat carriers (Dreher et al., 2009).

However, this interpretation must be qualified by the considerable uncertainty surrounding the effect of the DAT1/SLC6A3 VNTR on dopaminergic function. Both in vivo and in vitro studies have produced conflicting results, with some supporting greater DAT expression for the 9-repeat allele compared to the 10-repeat allele, while others report the opposite, or no relationship (Costa et al., 2011; Faraone et al., 2014). A recent meta-analysis of human imaging studies supports the first possibility when restricting the analysis to normal controls (Faraone et al., 2014). Disease status, development, and ancestry may all play a role in the functional consequences of DAT1 (Faraone et al., 2014; Franke et al., 2010; Shumay, Chen, Fowler, & Volkow, 2011). Even in the absence of changes in overall DAT expression, heterogeneities in DAT density and variations in neuronal morphology can substantially affect dopamine reuptake, which could contribute to the diversity of findings (Kaya et al., 2018).

It is also unclear the extent to which variation in DAT expression should be expected to influence tonic versus phasic DA. Phasic DA bursts are associated with salient stimuli

and have been shown to be associated with learning via reward prediction errors (Schultz, Dayan, & Montague, 1997; cf. Berridge, 2012). Various roles have been ascribed to tonic DA, including modulation of response vigor (Niv, Daw, Joel, & Dayan, 2007), exploration (Beeler et al., 2010), and the relative weighting of effort costs (Salamone, Correa, Farrar, & Mingote, 2007). DAT has a clear role in maintaining tonic DA concentrations (Efimova, Gainetdinov, Budygin, & Sotnikova, 2016; Sulzer et al., 2016). Accordingly, DAT has been attributed a major role in synaptic DA clearance after phasic release (Bilder et al., 2004), and pharmacological blockade of DAT alters DA transients and leads to long lasting increases in tonic DA (Floresco, West, Ash, Moore, & Grace, 2003; Ford, Gantz, Phillips, & Williams, 2010). However, detailed biophysical modeling suggests that diffusion is responsible for synaptic clearance of DA, with DAT having a (potentially limited) role in shaping the radius and duration at which DA bursts could activate receptors via volume transmission (Arbuthnott & Wickens, 2007; Cragg & Rice, 2004; Rice & Cragg, 2008). Notably, increasing burst firing of DA neurons in the ventral tegmental area does not cause tonic increases in extracellular DA in the nucleus accumbens without DAT blockade (Floresco et al., 2003). Tonic DA may also indirectly influence phasic activity, though the direction of this influence is complicated to determine; elevated tonic DA due to increased tonic DA neuron firing may augment the peak and duration of DA bursts (Dreyer, Herrik, Berg, & Hounsgaard, 2010), but tonic concentrations may also inhibit phasic DA via autoreceptor feedback mechanisms (Bilder et al., 2004).

The performance of patients with schizophrenia provides an interesting counterpoint to the combined effect of COMT and DAT. Opposite to the Met/Met:9-repeat carrier genotype, the pathology of schizophrenia includes hyperdopaminergic tone in striatum and hypodopaminergic tone in PFC (Brisch et al., 2014; da Silva Alves, Figuee, van Amelsvoort, Veltman, & de Haan, 2008; Grace & Gomes, 2018; Slifstein et al., 2015; Weinberger, Berman, & Daniel, 1992). Notably, patients with schizophrenia demonstrate reduced instructional bias on the PST (Doll et al., 2014). They also seem to rely less on putatively PFC-mediated processes in uninstructed learning, including reduced use of win-stay, lose-shift strategies and poorer performance on the easiest AB pair, potentially indicative of a reduced tendency to maximize or otherwise use rule-based strategies (Doll et al., 2014; Waltz, Frank, Robinson, & Gold, 2007; Waltz, Frank, Wiecki, & Gold, 2011). Though the elevated performance on AB in the Met/Met:9-repeat carrier group in the present study was not significant, it provides further evidence of opposite behavioral effects of opposite dopaminergic profiles.

Our findings of reduced flexibility with increasing ratio of PFC to striatal DA are also in accord with the effects of COMT and DAT1 on reversal learning. Compared to Met homozygotes, Val homozygotes show greater learning-rate adaptation around reversals, leading to improved performance (Krugel et al., 2009). Notably, Val homozygotes have more differentiated prediction error signals in striatal regions and greater learning-rate-dependent modulation of frontostriatal connectivity, suggestive of more adaptive prefrontal modulation of striatal RL (Krugel et al., 2009). On the other hand, the DAT1 9-repeat allele is associated with greater perseveration after a reversal (den Ouden et al.,

2013). It is interesting to note that this perseveration effect was explained by the 9-repeat allele conferring a more rapidly decreasing learning rate with increasing experience, which may be related to the decreased learning-rate modulation of COMT Met homozygotes. Direct comparison is difficult, however, as different computational models were used in the two studies. Importantly, while den Ouden and colleagues attributed their findings to more robust striatal learning of the previous reward contingencies, in the case of Met/Met:9-repeat carriers in the present study, their performance in the training phase cannot be due to greater ingraining of previous experience; the bias in the present case was due to instruction, not experience, was robustly evident in the first training block, and persisted throughout training.

4.2. Stimulation weakly increased instructional bias

In contrast to the genetic effects, the effect of tDCS on performance was far more limited. In accord with our hypothesis, anodal subjects demonstrated modestly more protracted learning away from the instructions during the early blocks of training. However, there was no effect of cathodal stimulation, and no effect of either stimulation condition during the test phase.

While the isolation of the effect to the training phase makes sense in light of the postulated division between frontal and striatal systems during training and test (Frank et al., 2007), it is at odds with the finding of increased bias at test associated with the COMT Met allele. It may be the case that genotypic effects on frontostriatal DA balance

or frontostriatal connectivity (discussed above) allow for greater biasing of striatum by PFC than is possible with single-session tDCS.

4.3. Mechanisms of instructional bias

The mechanisms underlying instructional bias are under debate. Proposals include models in which instructions bias striatal reward learning (learning bias models) (Biele et al., 2009; Doll et al., 2009) or those in which instructions affect choice rather than learning (choice bias models) (Doll et al., 2009). Evidence in favor of each of these classes of models has been mixed. Past computational modeling has tended to support learning bias models (Biele et al., 2009, 2011; Doll et al., 2011, 2009) but does not unequivocally rule out choice bias models (Doll et al., 2011, 2009). A number of neuroimaging studies have favored neither class of models, finding blunted activation in basal ganglia structures during instructed/prior knowledge conditions, suggesting a suppression of RL (Biele et al., 2011; Delgado et al., 2005; Fouragnan et al., 2013; Li et al., 2011). However, one study found overall decreased activity in reward structures but activity consistent with a learning bias in the form of an “outcome bonus” for choosing the instructed stimulus (Biele et al., 2011).

Adding to this debate, we find that our training phase results can be explained by a novel choice bias model—the decision bias model—containing a fixed bias for choosing the instructed symbol. This is in contrast to past work, which has found that a standard RL model without instructional bias best fits training phase performance, despite clear behavioral effects of instruction during training (Doll et al., 2011, 2009). Our model also

better predicted behavioral performance on CD trials compared to the Bayesian hypothesis testing model, a choice bias model previously shown to provide a reasonable fit to some subjects' training data and to be sensitive to effects of COMT (Doll et al., 2011, 2009). These results thus provide stronger evidence for the existence of a choice bias mechanism during training.

The decision bias and Bayesian hypothesis testing models differ in a number of regards (see section 2.7), with the most prominent differences being in the type of learner (standard Q-learning versus Bayesian Q-learning) and in the nature of the bias (fixed versus variable). We cannot say with certainty which of these factors most contributes to the superior performance of the decision bias model, though comparing our pattern of results to past work suggests that the Bayesian learner detracted from the performance of the model; all else equal, a variable bias should presumably better capture the behavior of a putative fixed bias agent than no bias. That said, an important direction for future work is to introduce a variable bias into the standard Q-learning framework and compare this to a fixed bias. This poses a challenge, since the uncertainty information used to implement adaptivity in the Bayesian framework is not present in the standard framework.

Again contrary to prior results, the decision bias model also best explained performance at test. While model comparison and striatal dopaminergic genetic effects have been previously taken as evidence of a learning bias mechanism at test (Doll et al., 2011, 2009), the supposition that the test phase primarily measures learning free of choice effects has recently come into question (Shiner et al., 2012; Smittenaar et al., 2012), in keeping with a broader role of DA in modulating motivation and learned value

representations (Berridge, 2012; Cagniard et al., 2006; Medic et al., 2014). Further supporting our finding, a recent reevaluation of test phase performance using an alternative model redescribed the learning bias for one striatal genotype as a choice bias (Collins & Frank, 2014). These discrepancies highlight the fact that model comparison results are dependent on the models tested. Additionally, in light of the evidence from other studies, there is no reason to think choice bias and learning bias mechanisms are mutually exclusive. However, the complexity of a model implementing both forms of bias would likely pose identifiability issues. We suggest that along with continued refinements to computational models, novel experimental designs capable of teasing apart these different possibilities will be necessary to advance our understanding of the mechanisms of instructional control.

4.4. Specificity of the effects and limitations

While there is good evidence that the expression of COMT and DAT1 are regionally specific, caution must be taken in interpreting the results of stimulation, as the lack of focality of tDCS prevents strong claims about effects on specific brain regions.

Stimulation could have altered the function of other brain areas involved in RL, including orbitofrontal cortex (O'Doherty, 2004). Neuroimaging and current modeling have even shown tDCS effects in subcortical structures, including the basal ganglia (Sadleir, Vannorsdall, Schretlen, & Gordon, 2010; Weber, Messing, Rao, Detre, & Thompson-Schill, 2014). However, the lack of stimulation effects on uninstructed learning and test phase performance somewhat militates against these possibilities.

Importantly, while our sample size was large for a tDCS study (Minarik et al., 2016) and was larger than the original report of the effects of COMT on instructed RL (Doll et al., 2011), these results should be replicated, particularly in light of the weakness of the tDCS effects and the small sample size of some genotypes. In the latter case, the low frequencies of the COMT Met and DAT1 9-repeat alleles in the population make collecting adequate samples of these groups challenging (Auton et al., 2015; Doucette-Stamm et al., 1995; Vandenberg et al., 1992). Because access to such samples is difficult outside of large cohort studies, we took statistical steps within our sample to ensure the robustness of our genetic results. Given the known interaction of COMT and task on the effects of prefrontal stimulation (Nieratschker, Kiefer, Giel, Krüger, & Plewnia, 2015; Plewnia et al., 2013), larger samples would also permit an examination of genotype x stimulation interactions. Though a between-subjects design was necessary in this study due to the use of deception, future examinations of this topic could also be improved by the development of within-subjects designs. Finally, it is conceivable that there is more opportunity to decrease bias than increase it, given the overwhelming feedback subjects receive in contradiction to the instructions. Unfortunately, cathodal tDCS, which could in principle be used to test this hypothesis, failed to elicit an effect in the present case and is demonstrably unreliable (Jacobson et al., 2012; Nozari et al., 2014). Future studies using theta-burst transcranial magnetic stimulation may be an appropriate alternative.

4.5. Conclusion

In sum, the present study provides further evidence for the role of PFC in biasing instructed RL, and additionally highlights the importance of frontostriatal DA balance in modulating top-down inputs. Such top-down regulation of learning by PFC is consistent with increased cognitive control leading to both costs and benefits (Chrysikou et al., 2014). Understanding the interplay of cognitive control and learning is thus key to establishing what level of control is most adaptive in a given situation. This endeavor will ultimately require delineating the relationship between computational and neurocognitive factors in learning and choice.

III. THE MODULATION OF BRAIN NETWORK INTEGRATION AND AROUSAL DURING EXPLORATION

1. Introduction

The brain has a remarkable capacity to adaptively shift processing to support a diverse array of behavioral goals, contextual demands, and environmental changes. This fact raises two fundamental questions: What neural mechanisms allow the brain to rapidly shift between states that form the substrates of different cognitive processes and behaviors, and how does the brain maintain a balance between the stability necessary to support ongoing behavior while maintaining the flexibility necessary to adapt to new exigencies? A number of theoretical proposals have pointed to a role for catecholamines in answering these questions, and in particular the neuromodulatory actions of norepinephrine (NE), a key component of physiological arousal (Arnsten, Paspalas, Gamo, Yang, & Wang, 2010; Aston-Jones & Cohen, 2005; Bouret & Sara, 2005; Yu & Dayan, 2005). The primary source of NE in the brain is the locus coeruleus (LC), a pontine nucleus that projects widely throughout the cortex (Berridge & Waterhouse, 2003a). NE has complex effects at single neuron level, but a common finding is that it increases the signal-to-noise ratio of neural responses, effectively modulating the gain of the neural response function (Berridge & Waterhouse, 2003a; Hasselmo, Linster, Patil, Ma, & Cekic, 1997; Hurley, Devilbiss, & Waterhouse, 2004), which simulations suggest can collectively lead to changes in functional connectivity and network topology (Eldar et al., 2013; Shine, Aburn, Breakspear, & Poldrack, 2018a). These features make the LC-NE system well situated to effect large-scale changes in brain networks and cognitive

function. Several prominent theories have ascribed this system just such a role, suggesting the LC-NE system resets functional brain networks in support of specific behaviors/cognitive states as dictated by environmental demands (Bouret & Sara, 2005), shifts the balance of information processing from top-down to bottom-up depending on the uncertainty of internal world models (Yu & Dayan, 2005), or shifts the brain between states of exploration and exploitation based on ongoing estimates of task utility (Aston-Jones & Cohen, 2005).

Recent studies have begun to explore the association between LC-NE activity and functional brain networks using human neuroimaging. Utilizing the fact that LC activity leads to increases in pupil diameter (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Joshi, Li, Kalwani, & Gold, 2016; Rajkowski, Kubiak, & Aston-Jones, 1993; Reimer et al., 2016; Varazzani, San-Galli, Gilardeau, & Bouret, 2015), studies have found that elevated pupil diameter is associated with stronger overall functional connectivity and greater clustering of functional connections (Eldar et al., 2013; van den Brink et al., 2016b; Warren et al., 2016), as well as an increase in the diversity of connectivity between functional communities, potentially indicating greater integration among brain networks (Shine et al., 2016). NE-linked changes in functional connectivity have also demonstrated spatial patterning consonant with specific catecholamine receptor distributions in humans (van den Brink, Nieuwenhuis, & Donner, 2018) and mice (Zerbi et al., 2019). Pharmacological manipulation of NE with Atomoxetine, a norepinephrine transporter (NET) blocker, has produced conflicting results, with resting-state studies finding decreased connectivity between networks (van den Brink et al., 2016b; see Guedj

et al., 2016 for a similar result in macaques), but increased connectivity between networks in a task-based study (Shine, van den Brink, Hernaus, Nieuwenhuis, & Poldrack, 2018b).

The heterogeneous results across studies likely stem from a number of factors, including differences in the methods used to construct and analyze brain networks, as well as differences in neural response between endogenous fluctuations of LC-NE activity and manipulation with Atomoxetine, which influences LC firing in addition to increasing cortical NE levels (Bari & Aston-Jones, 2013). Importantly, the divergence between task and rest effects may stem from the known inverted-U and state-dependent properties of the actions of catecholamines (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003a; McGinley et al., 2015; Robbins & Arnsten, 2009). Given that the actions of NE depend on the underlying state of the system, it is critical to ask what the relationship between brain network organization, LC-NE activity, and task performance is for particular classes of behaviors. To date, however, the relationship between NE and functional connectivity has not yet been assessed within the context of a task with an established relationship between NE-associated arousal and behavior.

The role of the LC-NE system in mediating between exploration and exploitation provides a strong place to begin to form these links. It has been proposed that increases in tonic LC-NE activity promote disengagement from the current task (exploitation) in order to seek alternatives (exploration) (Aston-Jones & Cohen, 2005). Direct LC stimulation promotes patch leaving and general disengagement during foraging (Kane et al., 2017), and pupil diameter has been found to increase with exploratory choice (Jepma &

Nieuwenhuis, 2011) and with decreases in task utility signaling the need to disengage from the current course of action (Gilzenrat et al., 2010). More broadly, elevated tonic LC activity and pupil diameter have been linked to distractibility (Aston-Jones & Cohen, 2005; Bouret & Sara, 2005; Ebitz & Platt, 2015; Unsworth & Robison, 2016; van den Brink, Murphy, & Nieuwenhuis, 2016a). Notably, a number of studies have suggested that task performance in cognitively demanding tasks is supported by increased integration among functional brain networks, with poorer performance predicted by decreased integration (Braun et al., 2015; Ekman, Derrfuss, Tittgemeyer, & Fiebach, 2012; Giessing, Thiel, Alexander-Bloch, Patel, & Bullmore, 2013; Shine et al., 2016; Vatansever, Menon, Manktelow, Sahakian, & Stamatakis, 2015). This suggests a potential parallel between elevated LC-NE activity and brain network integration—namely, that elevated LC-NE activity may lead to decreased functional integration, which may in turn provide a substrate for exploration.

We test this hypothesis in the present study. Subjects completed a two-armed bandit task while undergoing continuous fMRI and pupillometry. In order to meet the goal of linking arousal, functional connectivity, and behavior, we examined dynamic functional connectivity (Bassett et al., 2011, 2013; Fedorenko & Thompson-Schill, 2014), going beyond the static connectivity measures used in most prior studies in this domain to more tightly link arousal and connectivity changes, in the context of exploration. We also introduced a volatility manipulation between blocks to engender block-level differences in the rate of exploration. In keeping with our hypothesis, we predicted that exploration would be associated with increases in pupil diameter and decreases in brain network

integration. At the block level, we predicted that compared to low volatility blocks, high volatility blocks would be associated with greater exploration, greater pupil diameter, and lower integration.

2. Methods

2.1. Subjects

Forty-three subjects (24 female, $M_{\text{age}} = 23.28$ years) completed the study. Informed consent was obtained from each subject in accordance with the University of Pennsylvania IRB. All subjects in the final sample (1) were right-handed; (2) were between 18 and 35 years old; (3) had normal or corrected-to-normal vision; (4) had no known learning impairments or history of neurological or psychological disorders; and (5) were not currently taking any psychiatric medications or medications that are known to affect the autonomic nervous system. Three subjects were excluded due to technical difficulties at the scanner, and two subjects were excluded because it was later determined they did not meet the above inclusion criteria. Four additional subjects were excluded from the analyses for excessive head movement during scanning (average framewise displacement across runs > 0.2 mm), for a final sample of 34 (20 female, $M_{\text{age}} = 22.82$ years).

2.2. Materials and procedure

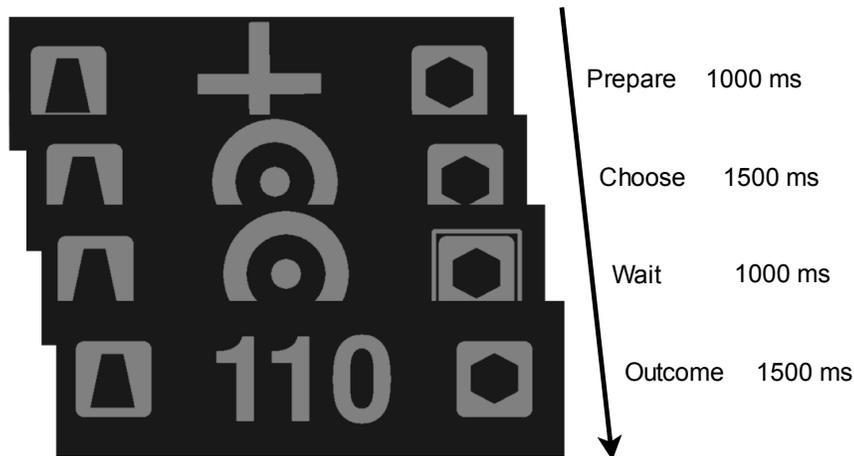


Figure 3.4. Stimuli and trial timing for the Leapfrog task. Each trial was followed by a one-second ITI during which a light gray rectangle was present in the center of the display to maintain luminance. Note that stimuli have been made higher contrast than they were during the experiment.

Subjects completed the Leapfrog bandit task (Knox, Otto, Stone, & Love, 2012). In this highly constrained two-armed bandit task (Figure 3.1), the options are always 10 points apart in value and when selected deliver payoffs deterministically. After every trial, with probability $P(\text{flip})$ the currently lesser-valued option may jump in value by 20 points to become the superior option. Which option is better thus alternates throughout the task, and subjects must balancing choosing the option that based on their current knowledge is the best (exploiting) with sampling the other option to find out if it has improved (exploring). The constrained nature of this task is advantageous among other reasons for the fact that trials can be classified as exploratory or exploitative solely on the basis of behavior, without recourse to model-based analyses necessitated by drifting bandits (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Ebitz, Albarran, & Moore, 2018).

Subjects completed four blocks of the task, with 80 trials per block (320 trials total). To minimize luminance-mediated changes in pupil diameter, task stimuli were luminance-matched grayscale images and were only modestly brighter than the background intensity. $P(\text{flip})$ was fixed within blocks but alternated across blocks [low volatility: $P(\text{flip}) = 0.05$; high volatility: $P(\text{flip}) = 0.20$], with the order of alternation counterbalanced across subjects. At the start of block 1, the left and right options were set to a value of 100 and 110, respectively. In a separate behavioral session prior to the scan session, subjects were instructed about the structure of the task (including the initial option values), performed 8 practice trials to familiarize themselves with the controls and the task display, and then performed an identical version of the task to the scanner version, excepting that the stimuli were not luminance controlled. While subjects received information about the volatility levels by completing the behavioral session, they were not told about the volatility manipulation. To minimize eye movements, subjects were instructed to fixate on the center of the task display at all times, except during the ITI, when they were told to keep their gaze within a 189x179 pixel light gray rectangle in the center of the display. Subjects made their responses with the index and middle finger of their right hand. Because the increase in payoffs throughout the task could distort choice behavior, subjects were incentivized to choose the currently best option on all trials rather than maximize their payoffs (Otto, Knox, Markman, & Love, 2014). Subjects were paid \$10/hr for the behavioral session (length 1 hr) and \$20/hr for the scan session (length 1.5–2 hrs) plus a bonus determined by $p \times \frac{b_{max}}{n_{trials}}$, rounded to the nearest dollar, where p is the number of choices of the currently best option, b_{max} is the

maximum possible bonus (\$10 behavioral, \$15 scan), and n_{trials} is the total number of trials.

The fMRI session began with eye tracker calibration, after which scans were run in the following order: resting state 1, Leapfrog block 1, B0, Leapfrog block 2, Leapfrog block 3, T1, Leapfrog block 4, resting state 2. Subjects were reminded of the current option values at the start of blocks 1, 2 and 4.

2.3. MRI data acquisition

Magnetic resonance images were collected using a Siemens Prisma 3T scanner (Siemens Medical Systems, Erlangen, Germany) with a 64-channel head coil. T1-weighted anatomical images were acquired (MPRAGE; repetition time [TR] = 1810 ms; echo time [TE]=3.45 ms; flip angle [FA]=9°; field of view [FOV]=240 mm; matrix = 256 X 256; voxel size = 0.9 X 0.9 X 1.0 mm²; 160 slices). During task runs, T2*-weighted functional volumes were collected using multiband echo planar imaging (EPI; TR = 1000 ms; TE = 30 ms; FA = 60°; FOV = 208 mm; matrix =104 X 104; voxel size = 2.0 X 2.0 X 2.0 mm²; 72 slices; multi-band acceleration factor = 6). We additionally collected resting state scans (not reported here; TR = 500 ms; TE = 25 ms; FA = 30°; FOV = 192 mm; matrix =64 X 64; voxel size = 3.0 X 3.0 X 3.0 mm²; 48 slices; multi-band acceleration factor = 6) A field map was also acquired for distortion correction of the EPI images (TR = 580 ms; TE 1 = 4.12 ms; TE 2 = 6.58 ms; flip angle = 45°; voxel size = 3.0 mm x 3.0 mm x 3.0 mm; FoV = 240 mm).

2.4. MRI preprocessing

Preprocessing was performed using FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) and FreeSurfer (Fischl, 2012). Cortical reconstruction and volumetric segmentation of the anatomical data was performed with FreeSurfer. Functional data were despiked by replacing values greater than 7 RMSE from a 1-degree polynomial fit to the time course of each voxel with the average value of the adjacent TRs. Motion correction parameters were computed by registering each volume of each run to the middle volume using a robust registration algorithm (`mri_robust_register`; Reuter, Rosas, & Fischl, 2010) and voxel shift maps for EPI distortion correction that were calculated using PRELUDE and FUGUE (Jenkinson, 2003, 2004); the resulting transformations were combined and simultaneously applied to the functional images. Boundary-based registration between structural and functional images was performed with `bbregister` (Greve & Fischl, 2009). To account for motion and physiological noise, the following nuisance time series were regressed from the functional data: (a) 24 motion regressors (Friston, Williams, Howard, Frackowiak, & Turner, 1996); (b) the five first principal components of non-neural sources of noise (i.e., white matter, CSF), obtained with FreeSurfer segmentation tools (`aCompCor`; Behzadi, Restom, Liau, & Liu, 2007); (c) cardiac and respiratory rhythms derived from pulse oximetry data collected during each scan (Verstynen & Deshpande, 2011); and (d) local noise, estimated as the average white matter signal within a 15 mm radius of each gray matter voxel (`ANATICOR`; Jo, Saad, Simmons, Milbury, & Cox, 2010). The data were then high-pass filtered with a cutoff frequency of 0.009 Hz.

2.5. Network construction

The cortex was parcellated into 200 regions based on the Schaeffer 200 parcel atlas (Schaefer et al., 2018). To this we added 15 subcortical regions segmented by FreeSurfer (Fischl et al., 2002). The average BOLD time series was extracted from each region, and functional connectivity between all pairs of regions was estimated via continuous wavelet coherence in the range of 0.08–0.125 Hz (Grinsted, Moore, & Jevrejeva, 2004). This frequency range has been previously shown to be sensitive to dynamic changes in task-based functional connectivity (Bassett et al., 2011; Braun et al., 2015; Gerraty et al., 2018; Sun, Miller, & D’Esposito, 2004). The continuous wavelet transform (CWT) was chosen over the more common discrete wavelet transform in order to provide additional sensitivity to time-varying changes around exploration. This procedure produces a connectivity value for each TR, sampled across the frequency range. Note that no windowing of the time series was performed prior to transformation, as the CWT is itself a sliding window method (i.e., a convolution), and additional windowing would produce unwanted edge effects (Grinsted et al., 2004). We then averaged across frequency to produce a single time-varying connectivity measure between each region. Finally, given that the resultant signal was heavily oversampled, the connectivity time series were then downsampled by a factor of 2 (final sampling rate of 0.5 Hz), yielding one 215 x 215 x 240 weighted adjacency matrix per task run.

2.6. Multislice Community Detection

In order to identify changes in network architecture over time, the connectivity matrices were submitted to a Louvain-like community detection algorithm (Mucha, Richardson, Macon, Porter, & Onnela, 2010) implemented in Matlab (Jeub, Bazzi, Jutla, & Mucha,

2011). This method, which has been used extensively to estimate time-varying community structure in functional brain networks, optimizes a multilayer quality function given by:

$$Q_{multislice} = \frac{1}{2\mu} \sum_{ijsr} \left[\left(A_{ijs} - \gamma_s \frac{k_{is}k_{js}}{2m_s} \delta_{sr} \right) + \delta_{ij} \omega_{jsr} \right] \delta(g_{is}, g_{jr}) \quad (1)$$

where the adjacency matrix of layer s has components A_{ijs} , g_{is} gives the community assignment of node i in layer s , g_{jr} gives the community assignment of node j in layer r , k_{js} is the intralayer strength of node j in layer s , $c_{js} = \sum_r \omega_{jsr}$ is the interlayer strength of node j in layer s , $\kappa_{js} = k_{js} + c_{js}$ is the strength of node j in layer s , and total edge weight of the network is given by $\mu = \frac{1}{2} \kappa_{jr}$. The quantity $\frac{k_{is}k_{js}}{2m_s}$ corresponds to the Newman-Girvan null model, where $m_s = \frac{1}{2} \sum_{ij} A_{ijs}$ is the total edge weight in layer s . The structural resolution parameter γ_s of layer s and the interlayer coupling parameter ω_{jsr} from node j in layer s to node j in layer r tune the size of the communities within each layer and the number of modules across layers (i.e., time), respectively. In this case, the structural resolution parameters were assumed to be constant across layers ($\gamma_s = \gamma$); the interlayer coupling parameters were set to a constant value ω where s and r were immediately adjacent layer and were set to 0 everywhere else, producing an ordered multilayer network.

The choice of γ and ω is not entirely straightforward. Often they are left at a default value of 1. In other instances, they are selected to optimize some quantity, such as $Q_{multislice}$ or other network measures of interest (Weir, Emmons, Gibson, Taylor, & Mucha, 2017).

Then, given the near degeneracy of the modularity landscape (Good, De Montjoye, & Clauset, 2010), the modularity maximization algorithm is run a number of times (e.g., 100) at the selected parameter values. To avoid dependence of our results on a particular point in parameter space and to increase sensitivity to fluctuations in integration regardless of scale, here we repeated the modularity maximization procedure a single time across a range of parameter values ($\gamma \in [1.14, 1.19]$ discretized by a step size of 0.01; $\omega \in [0.05, 0.85]$ discretized by a step size of 0.05) rather than multiple times at a single set of parameter values (see Vaiana, Goldberg, & Muldoon, 2019 for a related approach). The range of γ was chosen such that on average the number of non-singleton communities in a layer approximated the number of non-singleton cognitive systems in our resting-state reference partition (see below); the range of ω gamma was chosen to optimize network flexibility, which quantifies how often nodes switch communities across layers (Bassett et al., 2011).

Maximizing multilayer modularity is known to face a number of computational issues, particularly at lower values of ω . To mitigate these issues and improve the quality of the multilayer partitions, during each step of the Louvain algorithm, instead of choosing moves in an entirely greedy manner, moves were chosen probabilistically in proportion to their increase in the multilayer quality function (Bazzi et al., 2016; Jeub et al., 2011). Additionally, we used an iterated algorithm for maximizing modularity. After each run of the Louvain algorithm, community assignments were revised to maximize the persistence of communities across time without altering the intralayer community structure (Bazzi et al., 2016; Jeub et al., 2011). The resultant partition was then used as the starting point for

an additional run of the Louvain algorithm, and this procedure was repeated until the output partition converged (Jeub et al., 2011). These steps were repeated across the parameter grid, yielding 102 time-varying networks per run.

2.7. Integration

At each time point, a module allegiance matrix P^t was constructed, with entries:

$$P_{ij}^t = \frac{1}{O} \sum_{o=1}^O a_{ij}^{ot} \quad (2)$$

where O is the number of final output partitions (102) and the allegiance value a_{ij}^{ot} for nodes i and j is 1 if the nodes were placed in the same community at time t of partition o and 0 otherwise. Intuitively, P_{ij}^t is the probability that two nodes were placed in the same community at a given time point, across the parameter space (see Braun et al., 2015 for a similar approach to computing network measures per time window).

In order to then use the modular allegiance matrices to assess the interaction between brain regions across time, we assigned each network node to a resting-state cognitive system. All cortical nodes were previously assigned to one of seven resting-state systems identified from large-scale resting-state data (Schaefer et al., 2018; Yeo et al., 2011). All subcortical nodes were assigned to an eighth Subcortical system with the following exceptions: bilateral amygdala and hippocampus were placed in the Limbic system, while the brainstem was assigned to its own singleton system.

The integration of a brain region i in cognitive system s at time t can then be computed as:

$$I_i^{ts} = \frac{1}{N - n_s} \sum_{j \notin s} P_{ij}^t \quad (3)$$

where N is the total number of nodes (brain regions) and n_s is the number of nodes in system s (Mattar, Cole, Thompson-Schill, & Bassett, 2015). Integration thus quantifies the probability at a given time that a node from a given cognitive system is placed into the same community as nodes from other cognitive systems. Averaging integration across nodes then provides a measure of the global level of integration in the brain at each time.

Integration can also be computed at the system and between-system levels. The integration of a system s with the rest of the brain (i.e., all systems not s) is:

$$I_s^t = \frac{1}{n_s(N - n_s)} \sum_{i \in s} \sum_{j \notin s} P_{ij}^t \quad (4)$$

indicating the tendency for nodes from system s to be placed into communities with nodes from other systems at time t . Similarly, the integration between two systems k and l is given by:

$$I_{kl}^t = \frac{1}{n_k n_l} \sum_{i \in k} \sum_{j \in l} P_{ij}^t \quad (5)$$

where n_k is the number of nodes in system k and n_l is the number of nodes in system l .

High integration between two systems at a given time indicates a departure from resting-state network structure and is suggestive of strong functional interactions between

cognitive systems. All of these integration measures can be further averaged across time to provide block-wise measures of global, system, and between-system integration.

2.7.1. *Peri-explore integration analysis*

Statistical analysis of change in the integration time course around exploration presents a number of methodological challenges. The time series is strongly autocorrelated, which increases the risk of type I error due to violation of the independence assumption of linear regression. The response to exploration is of an unknown functional form and potentially non-monotonic, making standard linear regression—even using polynomial terms—a potentially poor fit. Finally, unlike in the pupil analyses, there is no clear contrast or baseline, so tests of $H_0 = 0$ at each time point are not necessarily appropriate either.

To address all of these issues, we utilized generalized additive mixed models (GAMMs) in the peri-explore integration analyses. GAMMs are an extension of the regression framework that allow for the fitting of arbitrary (e.g., nonlinear, nonmonotonic) functions, including both linear and nonlinear random effects terms (Wood, 2017). These nonlinear functions, or *smoother*s, are fit using maximum likelihood estimation using a weighted sum of basis functions. The basis functions are selected from families of penalized splines, where overfitting is mitigated and therefore smoothness is enforced by a “wiggleness” penalty on basis function coefficients. The appropriate smoothness for a given data set is controlled via smoothing parameters that are estimated as part of the fitting procedure (See Baayen, Vasishth, Kliegl, & Bates, 2017; Pedersen, Miller,

Simpson, & Ross, 2019; and van Rij, Hendriks, van Rijn, Baayen, & Wood, 2019 for tutorials, and Wood, 2017 for additional technical and mathematical details).

Prior to model fitting, peri-explore integrations time courses were extracted and processed as follows. After identifying the time points in the integration time course that contained each exploratory choice, we extracted the time series immediately preceding (following) the choice window, up to the previous (next) exploratory choice. In order to isolate the effect of a single exploratory choice given the sluggishness of the integration time course, we restricted the analysis to explore choices preceded by a minimum of 2 exploit trials and followed by a minimum of 4 exploit trials. We additionally excluded the first and last peri-explore periods of every block. We also did not count as exploration trials in which subjects explored immediately following a missed flip (i.e., subjects exploited and saw a change). The final analysis window was then restricted to encompass the 12 s prior to the explore window extending to 18 s post-explore. We then downsampled the time series to 0.25 Hz as a first step in mitigating autocorrelation.

All GAMMs included a smooth for Time and by subject random smooths for Time. Models also included by time course linear random intercepts and slopes in order to account for additional variance due to drifts in integration over time, which helps to further alleviate autocorrelation in the residuals (van Rij et al., 2019). Because model residuals were still autocorrelated, we also introduced an AR1 model to each GAMM. For analyses of global integration, the AR1 parameter that minimized AIC in a grid search ($\rho \in [0.00, 0.99]$ in steps of 0.01) was selected for the final model (Wood, 2017). For by system integration, residual autocorrelation was very similar in each system, and

so we selected the ρ that minimized AIC for the model with the median AR1 value. The same approach was used for between system integration.

To confirm the results of the global integration GAMM, a permutation analysis was conducted. Within each block, the assignment of exploration time points to the integration time course was permuted 500 times, with the constraint that the distribution of inter-explore intervals remain constant across permutations. Peri-explore time courses were then extracted and analyzed as above, resulting in a GAMM fit for each permutation. The significance of the true data was then assessed relative to this distribution. Note that we took the somewhat unusual step of constructing our permutation distribution from the p -values of the smooths rather than the F values due to the fact that unlike in a standard parametric linear analysis, the number of degrees of freedom differs between models due primarily to differences in the wiggleness of the fit, and also due to slight differences in the amount of data in each permutation as a result of preprocessing exclusions. Using F values can thus produce conservative results, as smooths with fewer effective degrees of freedom may benefit from larger F values. Because the p -value computation takes degrees of freedom into account (Wood, 2013), it is thus a more appropriate measure in this case.

2.8. Additional network measures

To better characterize the network dynamics surrounding exploration, we computed a number of additional network measures, using the Brain Connectivity Toolbox (Rubinov & Sporns, 2010).

The average *strength*, s , of node i at time t was computed as:

$$s_i^t = \frac{1}{N-1} \sum_j A_{ij}^t \quad (6)$$

By averaging node strength separately for within and between system connections across the whole brain, *system segregation* (Chan, Park, Savalia, Petersen, & Wig, 2014) was computed as:

$$\text{system segregation} = \frac{\bar{s}_w - \bar{s}_b}{\bar{s}_w} \quad (7)$$

Unless otherwise noted, we computed system segregation relative to the Yeo cognitive systems, to match our procedure for integration, rather than to the module assignments at each time point.

The single-layer modularity Q (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) was computed at each time t using as input the module assignments derived from each run o of multilayer modularity (equation 1). Specifically

$$Q_o^t = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (8)$$

where $k_i = \sum_j A_{ij}$ and $m = \frac{1}{2} \sum_{ij} A_{ij}$ and o and t super/subscripts are omitted for clarity.

Q values were then averaged over o to produce a single Q_t at each time point. Finally, the *number of modules* was defined as the average number of modules present at each time point, averaged over runs of GenLouvain.

As with integration, significance of peri-explore modulation was assessed in the GAMM framework. Because of heavy skew in the data, the number of modules model was fit with an inverse gaussian regression (log link). A single AR1 parameter ρ was used for all strength-based measures (strength, system segregation) and all modularity-based measures (Q , number of modules).

2.9. Pupillometry

Eye position and pupil diameter of the right eye were recorded during scanning at a sampling rate of 250 Hz with an EyeLink 1000 Plus (SR Research) equipped with the Long Range Mount. Period of missing data due to blinks or other artifacts were linearly interpolated after removing an additional 25 samples (100 ms) surrounding the blink on either side. Additional artifacts were identified by computing the difference between consecutive samples of the pupil time course. High velocity periods, defined as samples differing in diameter by more than 50 in absolute value (a.u.) from the preceding sample were removed, and for runs of high velocity > 4 samples we additionally removed 25 samples on either side of the run, identical to the procedure described for blinks. These censored periods were then linearly interpolated. The pupil time course was then lowpass filtered with a 4 Hz cutoff. The data were then normalized by z-scoring within-subject across data from all functional runs. Gaze position data for time points missing or removed from the pupil time course were also interpolated. Blocks in which $> 50\%$ of the pupil data were missing or censored were not included in the analysis (two blocks from one subject).

2.9.1. Pupil analysis

Baseline pupil diameter was calculated as the average diameter in the last 500 ms of the fixation period at trial start. Pupil dilation was quantified as the maximal dilation in the 2.5 s between the beginning of the choice window and the presentation of the outcome. For trial-level analyses, data were downsampled to 50 Hz, and all models included gaze position as covariates. Analyses of the choice period also controlled for baseline pupil diameter at the start of the trial. Analyses for the outcome period instead controlled for average pupil diameter in the last 250 ms of the gap between the end of the choice window and the onset of the outcome stimulus.

For the post-explore pupil analysis, pupil diameter was downsampled to 2 Hz, since the focus was on slower changes in diameter over a longer time scale. We used the same restrictions on the data submitted to this analysis as described above for integration, except we relaxed the minimum number of exploit trials post-explore to 2. For analyses of the post-explore peak/minimum, we identified peaks as the maximum dilation in the period from 0–12 s post-explore. The post-peak minimum was then identified in the period from the peak to 18 s post-explore.

2.10. Pupil–network relationships

To characterize the relationship between pupil-linked arousal and integration, we first downsampled pupil diameter to the sampling rate of the TR and then applied a low-pass filter by convolving it with a gaussian with a standard deviation equal to the median wavelet scale used to compute wavelet coherence for the network analysis (9.80 s).

Finally, we downsampled the filtered time course to sampling rate of the integration time course (0.5 Hz). We then computed the cross-correlation between the pupil diameter and each network measure over the peri-explore period, using the same peri-explore criteria described above for peri-explore integration. To plot the cross-correlation and compute within- and across-subject averages, we first Fisher z -transformed the correlations.

Because the presence of autocorrelation biases the variance of sample correlations, we first corrected the z -transformed correlations for this bias, using the method of Pyper and Peterman (1998), producing Z -scores (Afyouni, Smith, & Nichols, 2019). This procedure effectively weights each z value in proportion to its effective degrees of freedom. We then averaged the Z -scores within subject and assessed the significance of the correlation at the peak lag using a one-sample t -test against 0.

2.11. Data analysis

Statistical analyses were performed in R (R Core Team, 2019). Linear and logistic mixed effects models were implemented in the lme4 package (Bates et al., 2015b), except when an AR1 model was fit for the residuals, in which case nlme was used (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019). Where possible, models included random intercepts for subjects and random slopes for all within-subjects variables (i.e., the maximal model; Barr, Levy, Scheepers, & Tily, 2013). In cases where the maximal model failed to converge or produced singular fits, we iteratively reduced the random effects structure until convergence, following steps outlined by Bates and colleagues (Bates, Kliegl, Vasishth, & Baayen, 2015a). Post-hoc comparisons were computed using the emmeans package (Lenth, 2016). GAMMs for the analysis of peri-explore integration

time courses were implemented in the mgcv package (Wood, 2017). Where noted, significance levels were corrected for multiple comparisons using the Bonferroni-Holm method. All block-level analyses and all analyses comparing volatility conditions discarded the first 10 trials of every block, in order to give subjects some time to adjust to the current volatility level.

3. Results

We first characterize the pupil response to exploration. We then examine the dynamic modulation of integration around exploration and relate changes in pupil diameter to changes in integration. Finally, we examine effects of the volatility manipulation on arousal and integration.

3.1. Exploration modulates pupil diameter

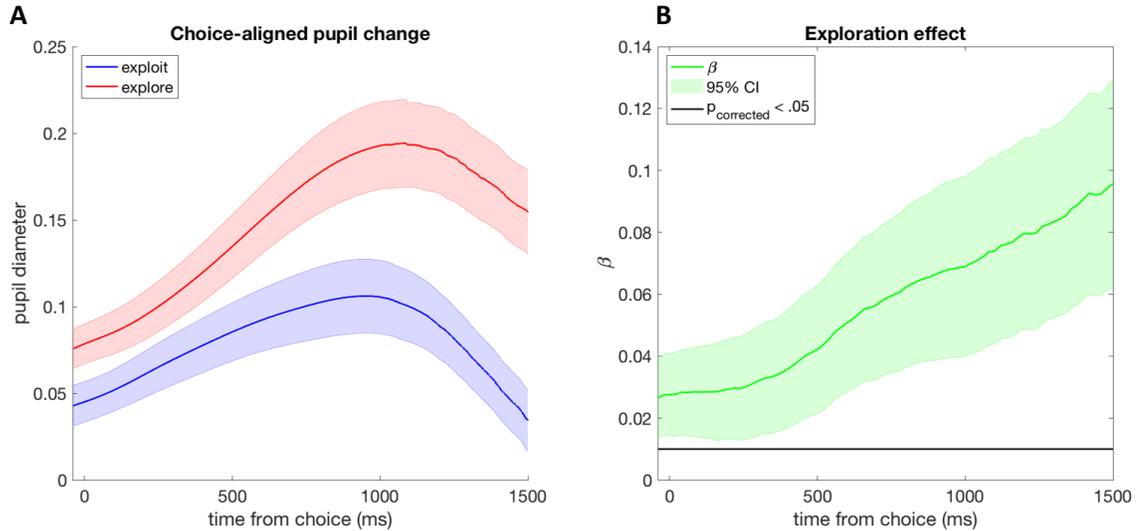


Figure 3.5. Pupil diameter is reliably modulated by choice type. **A** Average pupil response to explore choices and exploit choices across subjects. Pupil diameter is z-scored within subject, and the evoked response is calculated relative to a pre-trial baseline taken from the average in the 500 ms prior to the choice period. Here and throughout, error bars accompanying averaged data reflect the standard error of the mean (SEM). **B** The contrast of explore > exploit from a mixed-effects regression model for every time point. Pupil diameter was downsampled from 250 to 50 Hz. The regression model controls for baseline pupil diameter and gaze position.

Confirming our hypothesis, pupil diameter was reliably higher throughout the entirety of the choice period when subjects made explore choices compared to exploit choices (Figure 3.2; all $p_{\text{corrected}} < .011$). This difference was present at least 40 ms prior to the button press. Prior reports have found elevated baseline pupil diameter prior to exploration, distraction, and disengagement (Ebitz & Platt, 2015; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011). Given this, we also examined the pre-explore period (Figure 3.4A). Baseline pupil diameter varied significantly among the three trials just

prior to and including the explore trial ($F(2, 5032) = 6.56, p = .014$). This was driven primarily by a decrease in pupil diameter from the second to the first trial pre-explore ($\beta = -0.06, t(5032) = -3.62, p_{\text{corrected}} = .009$), potentially reflecting at least in part the diminishing influence of the previous exploratory choice. Though pupil diameter rose on the explore trial relative to the immediately preceding trial, this was not significant ($\beta = 0.03, t(5032) = 1.80, p_{\text{corrected}} = .14$), and the baseline diameter on the explore trial was still numerically smaller than that of two trials previous ($\beta = -0.03, t(5032) = -1.81, p_{\text{corrected}} = .14$). This suggests that in this task, the elevated pupil diameter post-explore was driven by the explore choice itself and not by prior ramping of arousal.

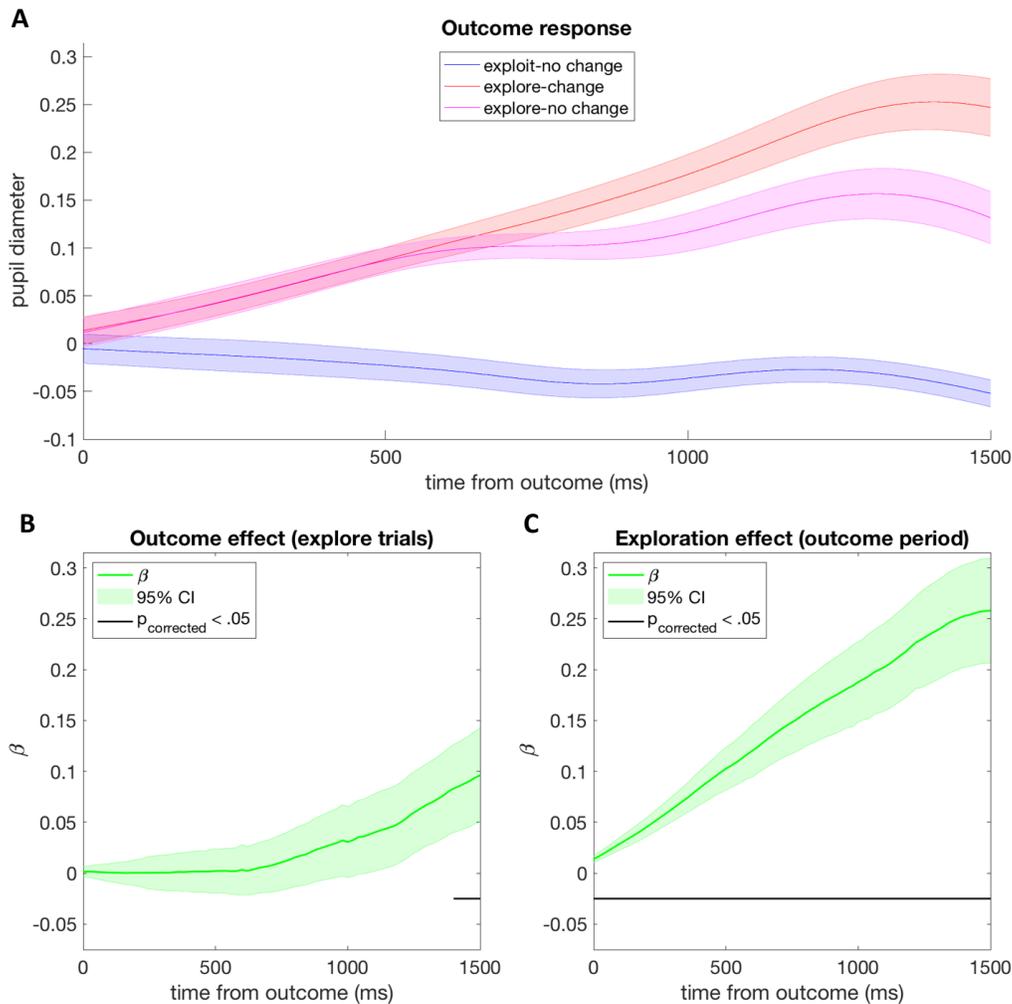


Figure 3.6. The effect of outcomes on pupil diameter. **A** Average pupil response to outcomes, separate by whether the choice was explore or exploit. The evoked response is calculated relative to the average pupil diameter in the 250 ms prior to presentation of the payoff. Note that exploit-change trials are not shown, as they were rare outcomes and were thus not analyzed. **B** The contrast of explore-change > explore-no change trials from a mixed effects model of the outcome period. Changes induce a reliable increase in pupil diameter at the end of the outcome period. **C** The contrast of explore > exploit-no change from the same model. This is the effect of exploration over and above the effect

during the choice period, as this model controls for average pupil diameter in the 250 ms prior to outcome presentation. The model also controls for gaze position.

Because pupil diameter is also modulated by outcomes, particularly if they are surprising (Alamia, VanRullen, Pasqualotto, Mouraux, & Zenon, 2019; Friedman, Hakerem, Sutton, & Fleiss, 1973; Lavín, San Martín, & Rosales Jubal, 2014; Nassar et al., 2012; Preuschoff, 't Hart, & Einhäuser, 2011), we also examined pupil dilation in response to changes in payoffs. In the Leapfrog task, because payoffs are deterministic excepting the stochastic jumps, outcomes will either be the same as when the option was last checked, or they will have jumped in value. Therefore, we divided trials into three classes, based on whether subjects explored and the payoff increased (explore–change), explored and the payoff was unchanged (explore–no change), or exploited and the payoff was unchanged (exploit–no change). Trials in which subjects exploited and the payoff increased (exploit–change) were excluded from the analysis as there were very few per subject. Given this, we contrasted the response to change within explore trials only.

Pupil diameter was slightly elevated in response to a change in outcome (Figure 3.3A,B). This separation began to emerge in the averaged data around 500 ms post outcome presentation but was only reliable in the last 100 ms of the outcome period (all $p_{\text{corrected}} < .047$). This effect was much smaller in magnitude than the continued effect of exploration on the pupil response (contrast of explore trials with exploit–no change trials), which was reliable throughout the outcome period (Figure 3.3A,C; all $p_{\text{corrected}} < 0.0001$). Note that this effect is not simply due to carryover from the choice period, as these analyses

controlled for pre-outcome pupil diameter; rather, this appears to reflect an extended influence of exploration on post-choice arousal.

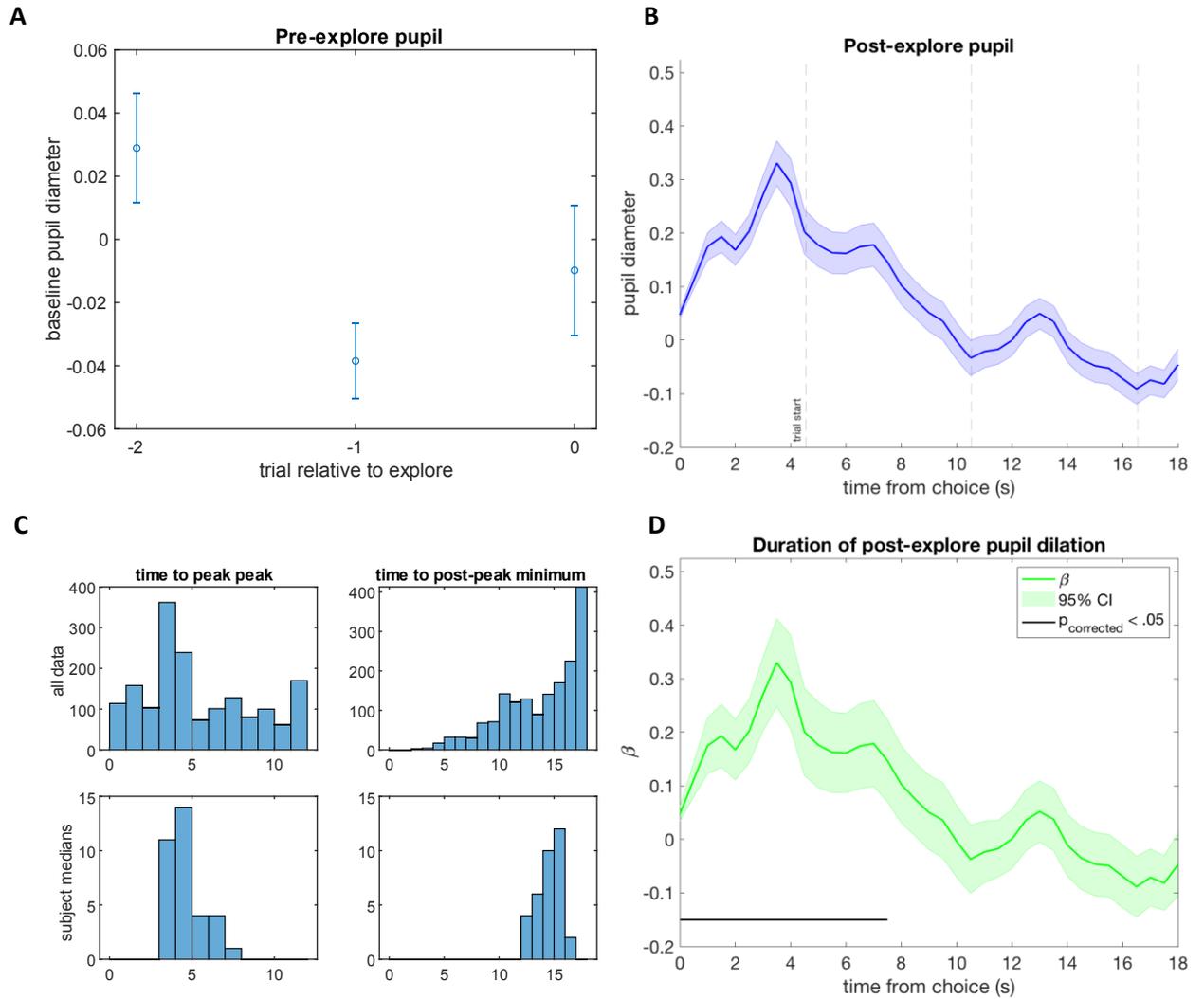


Figure 3.7. Modulation of pupil diameter pre- and post-explore. **A** Pre-explore baseline pupil diameter on the trials preceding exploration. Only the decrease from the second to the first trial pre-explore was significant. **B** The post-explore pupil time course, aligned to the explore choice. Dashed vertical lines indicate the approximate start times of subsequent trials. The small upward modulations in the time course shortly after each trial start are due to subsequent exploit choices. **C** The post-explore pupil diameter

latency to peak and latency from peak to the post-peak minimum (max 18s post-explore) across all data (top); the median latency to peak and post-peak minimum for each subject (bottom). **D** Regression model of the post-explore time course. Pupil diameter is significantly elevated above the explore trial baseline for 7.5 s post-choice.

We next sought to characterize the duration of the arousal response (Figure 3.4B–D). Pupil diameter was significantly elevated above the explore-trial baseline for 7.5 s post-choice, approximately the start of the outcome period of the subsequent trial (all $p_{\text{corrected}} < .015$). This result held when controlling for gaze position (all $p_{\text{corrected}} < 0.039$) and when additionally constraining the analysis to those epochs with minimal eye movements (< 50 pixels root mean squared; all $p_{\text{corrected}} < .028$). The sustained duration of the effect also does not appear to be primarily attributable to an artifact of averaging over subjects with variable exploration responses (Figure 3.4C). The median peak exploration response (median of within-subject medians) from 0–12 s post-explore occurred 4.0 s post-choice, which is very similar timing to the peak at 3.5 s in the time-averaged data (Figure 3.4B). Furthermore, the majority of individual subjects' median peaks were not significantly different from the group median (31/34 subjects, sign test [corrected]). Similarly, the median minimum pupil dilation in the window from the post-explore peak to 18 s post-explore was 14.5 s, identical to the time-averaged minimum (Figure 3.4C). This was consistent with the minimum in all subjects (34/34 subjects, sign test [corrected]). Nor was this time course significantly modulated by outcome type, though there was a small modulation that was significant at an uncorrected $p < .05$ level from 4–5.5 s post-choice, consistent with the effect seen at the trial level at the end of the outcome period and extending into the ITI and the start of the subsequent trial (Figure S1). The smearing out

of the outcome effect by time-locking on choice, as well as the trial restrictions imposed on this analysis, may have made it more difficult to detect the small modulation by outcome found in the trial-level data.

The elevation of pupil diameter with exploratory choice thus seems best explained as a transient increase in tonic arousal driven purely by the choice to shift from exploitation to exploration, rather than an artifact or a response to the outcome. Nor does increased arousal appear to be the cause of the exploratory choice, rather than its effect. However, one additional possibility is that pupil diameter is elevated in response to exploration due to the greater uncertainty in the outcome on explore trials as compared to exploit trials. Indeed, the probability of observing a change in option value on explore trials is fairly uncertain ($P(\text{change} \mid \text{explore}) = 0.41$), while it is very unlikely on exploit trials ($P(\text{change} \mid \text{exploit}) = 0.13$). If this were the case, it might be expected that the pupillary response to exploration would differ between volatility conditions, as $P(\text{change} \mid \text{explore})$ was higher in the high volatility blocks ($P(\text{change} \mid \text{explore,high}) = 0.57$; $P(\text{change} \mid \text{explore,low}) = 0.24$). This was not the case. There was no effect of volatility condition, nor any volatility x choice type interaction during the choice period (Figure S2; all $p_{\text{corrected}} = 1$). Similarly, there was no effect of volatility condition on the post-explore time course; though the high volatility condition demonstrated a slightly lower pupillary response in the first second post-choice, this did not survive correction for multiple comparisons (Figure S2; all $p_{\text{corrected}} > .62$). Given subjects' overall weak sensitivity to the volatility conditions, these results do not completely rule out a role for uncertainty in

driving choice effects, but they raise the possibility that exploratory choice itself, isolated from effects of uncertainty or surprise, can drive shifts in arousal.

3.2. Exploration transiently modulates peri-explore integration

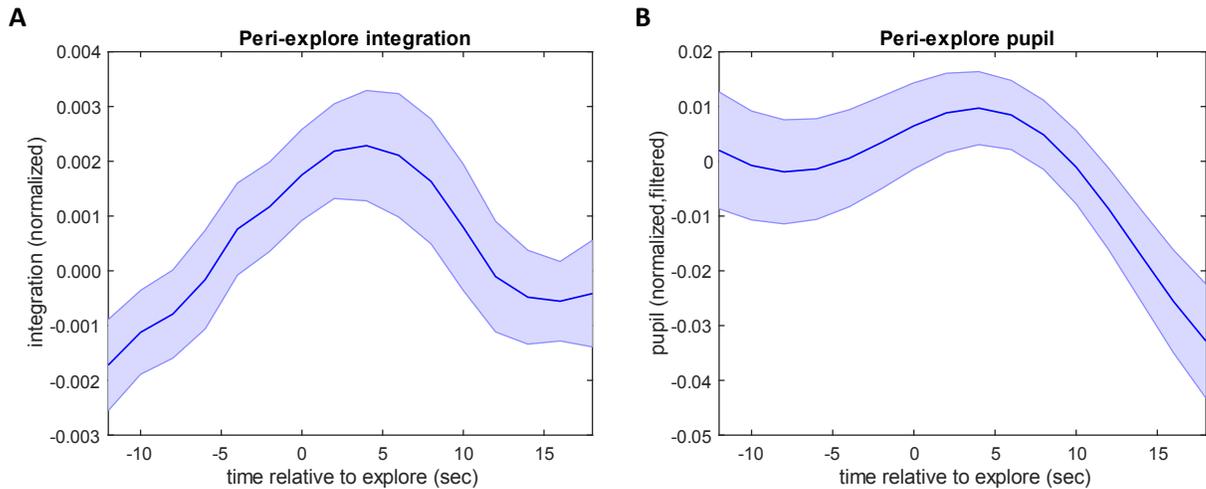


Figure 3.8. **A** The peri-explore integration time course is significantly modulated around exploration. All peri-explore time courses both here and below were mean-centered prior to averaging for display purposes. Uncentered time courses were used in the statistical analyses, and trial-to-trial variability was captured using by trial random effects. **B** The peri-explore pupil time course, downsampled to the sampling rate of the integration time course and low-pass filtered.

Integration was also significantly modulated around exploration (Figure 3.5A.; $F(3.32, 4551.90) = 4.03, p = .002$). Integration appears to increase leading up to exploration, peak around the explore choice, and fall thereafter. To rule out the possibility that this result was reflective of some more general oscillation in the data, we refit the GAMM on data in which the location of explore trials was permuted within each block (500 permutations). This analysis strongly suggested that the modulation was unique to exploration ($p = .008$).

To understand the factors driving this change in integration, it is important to answer two questions: 1) Which cognitive systems and their interactions contribute most to these dynamics? and 2) How do changes in integration relate to other global network properties?

3.3. Integration is modulated differentially across cognitive systems

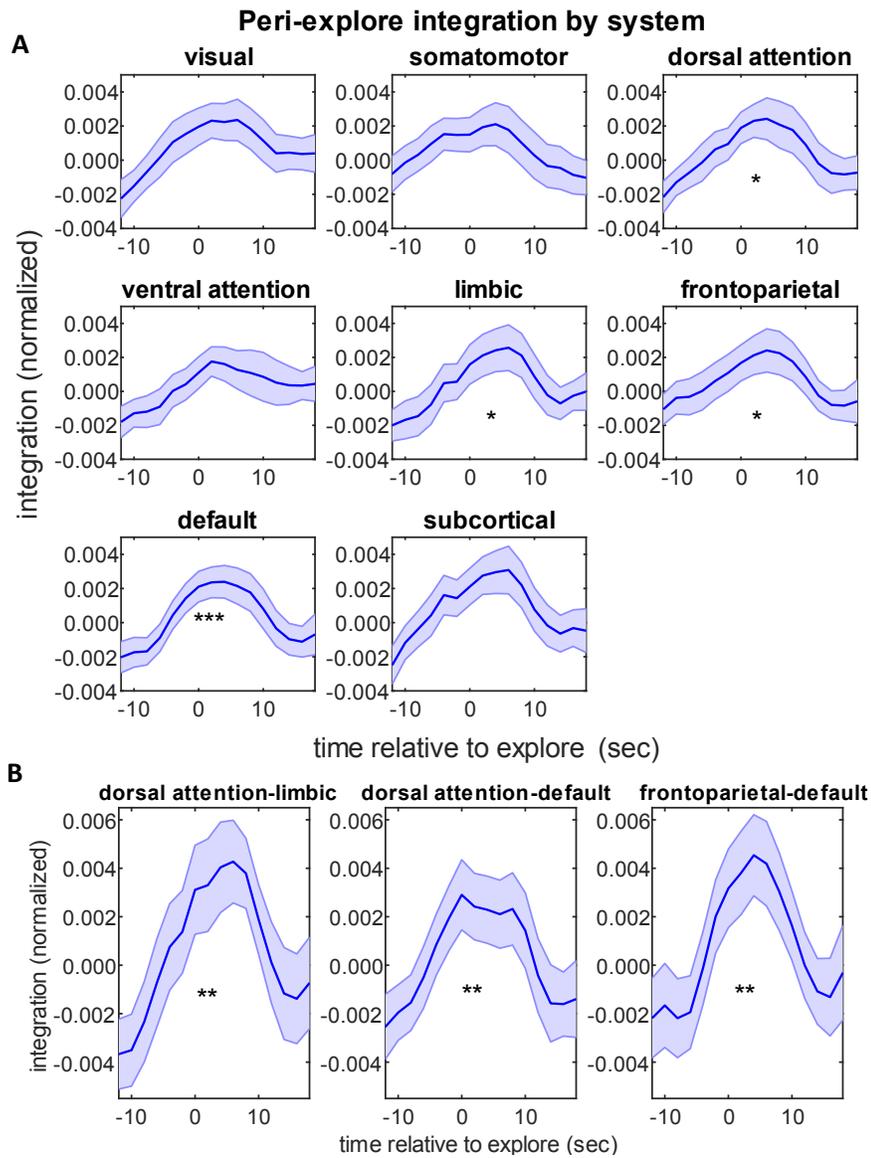


Figure 3.9. The modulation of peri-explore integration varies by cognitive system. **A** The integration of each cognitive system with all other systems (i.e., the rest of the brain). **B** Pairwise interactions between cognitive systems that demonstrated a significant modulation around exploration. * $p < .05$; ** $p < .01$; *** $p < .001$.

To answer the first question, we computed system-level integration, the integration of each cognitive system with all other systems (i.e., with the rest of the brain; see Methods). While qualitatively there was some evidence of the global modulation when examining each cognitive system individually, this was only significant for the dorsal attention, default, frontoparietal, and limbic systems (Figure 3.6A.; dorsal attention: $F(3.19, 4522.75) = 3.97$, $p_{\text{corrected}} = .02$; limbic: $F(3.34, 4650.21) = 3.50$, $p_{\text{corrected}} = .037$; frontoparietal: $F(3.32, 4576.66) = 3.36$, $p_{\text{corrected}} = .038$; default: $F(4.08, 4570.32) = 5.27$, $p_{\text{corrected}} = .0006$).

We then asked whether any interactions between cognitive systems differentially contributed to the system-level changes by computing between-system integration, the integration of two cognitive systems with each other. Significant modulation of between system integration was found only for dorsal attention–limbic, dorsal attention–default, and frontoparietal–default interactions (Figure 3.6B; dorsal attention–limbic: $F(3.36, 4466.03) = 5.36$, $p_{\text{corrected}} = .006$; dorsal attention–default: $F(3.57, 4428.86) = 5.11$, $p_{\text{corrected}} = .007$; frontoparietal–default: $F(4.03, 4368.86) = 5.19$, $p_{\text{corrected}} = .003$; see Figure S3 for all between-system integration time courses). Therefore it is not the case that integration was modulated uniformly throughout the brain, as might be expected under some theories of LC function (Eldar et al., 2013). Rather, changes in integration

demonstrated specificity, perhaps reflective of interactions between these systems underlying decisions to explore, or of changes in interactions between these systems providing the substrate for exploratory states.

3.4. Exploration induces complex changes in connectivity and topology

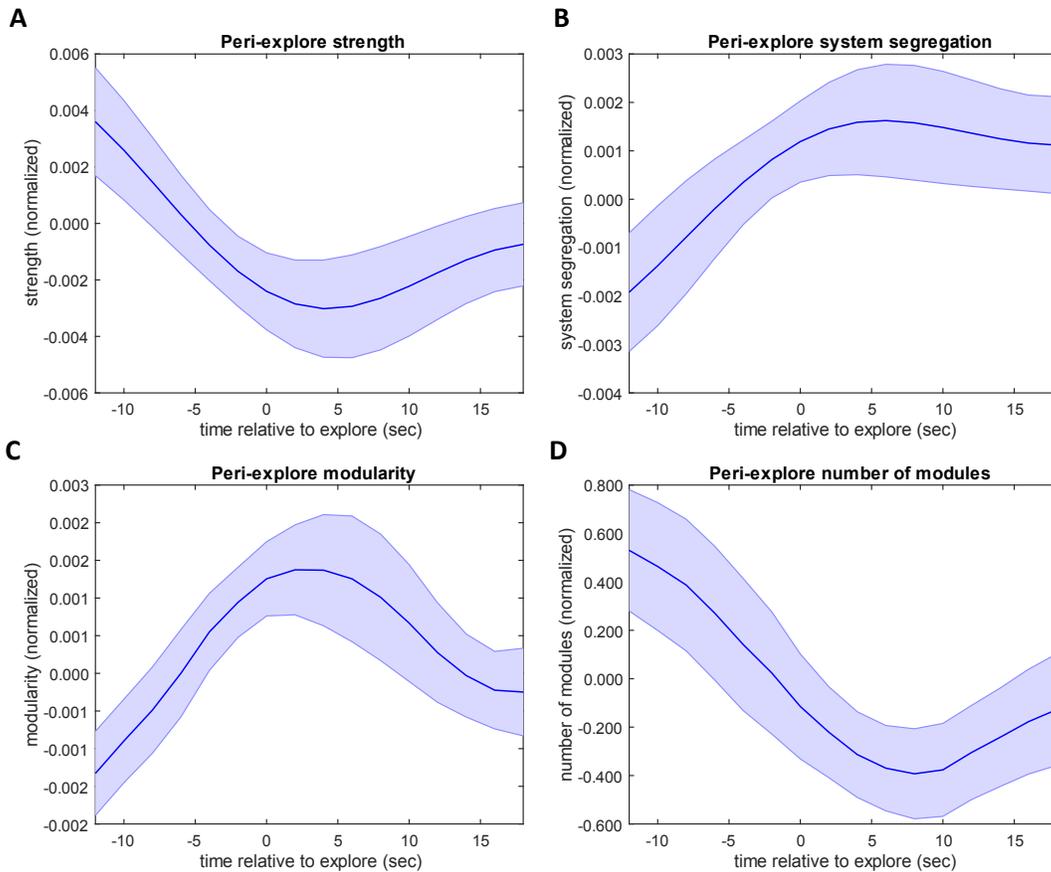


Figure 3.10. Average node strength, system segregation, modularity, and number of modules all show significant modulations in the peri-explore period.

Regarding the second question above, changes in integration between cognitive systems could be driven by multiple facets of the underlying connectivity and topology. For example, though integration is based on network topology and not directly on

connectivity, intuitively increases in integration might reflect a shift toward increased functional connectivity strength. Contrary to this expectation, average node strength demonstrated an opposing profile to integration, reaching a minimum and plateauing close to the time of choice (Figure 3.7A; $F(3.80, 4297.55) = 7.64, p < .0001$). To assess whether strength changed differentially within and between cognitive systems, potentially contributing to the change in integration, we computed a strength-based measure of system segregation—the difference in within versus between system connectivity, as a percentage of within system connectivity (see Methods). Thus, increases in this quantity reflect an increase in the relative strength of within-system connectivity. While both within and between system connectivity demonstrated a qualitatively similar peri-explore profile (Figure S4), system segregation demonstrated a positive modulation in favor of within system connectivity (Figure 3.7B.; $F(3.18, 4337.78) = 4.79, p = .0007$). This result was not driven by a mismatch between the assignment of nodes to cognitive systems relative to the dynamic modular structure of the network, as a similar pattern obtained when computing system segregation relative to the module assignment at every time point (Figure S4; $F(3.38, 4449.78) = 5.24, p = .0002$).

This increase in system segregation, usually inferred to reflect a *decrease* in the integration of network communities, suggests that the positive modulation of integration may rather reflect a transient topological shift toward fewer modules. This was indeed the case (Figure 3.7D; $F(4.38, 4268.60) = 5.27, p < .0001$). Finally, we asked how these changes in connectivity and topology related to the (single-layer) modularity of the network (see Methods), which is also often considered a measure of segregation

(Rubinov & Sporns, 2010). Because modularity is a measure of the extent to which intra-module strength is greater than expected, it might be expected to be positively associated with system segregation. Alternatively, it could be expected to track with the number of modules, as fewer modules often indicates a less modular structure. Here, we found that modularity demonstrated a positive fluctuation during the peri-explore period, in line with the increase in system segregation (Figure 3.7C; $F(3.68, 4522.20) = 6.10, p < .0001$).

In sum, around exploration, there is a temporary shift toward a smaller collection of more loosely connected modules that include nodes from a greater diversity of cognitive systems. This counterintuitively leads to an increase in measures normally taken to measure segregation (modularity, system segregation), while at the same time increasing our measure of integration. While these results are consistent with our hypothesis that integration would be modulated around exploration, they are not entirely in line with the directionality of the hypothesis—that exploration would decrease integration. This inconsistency is due both to the heterogeneity across measures and to the fact that the integration results could be consistent with either a localized peak concomitant with exploration, or with an increase during exploitation followed by a decrease following exploration. Unfortunately, the temporal resolution of this analysis is not sufficient to fully disentangle these possibilities. Notably, using wavelet analysis, the minimum size of an effect produced by a transient will be approximately the size of the wavelet’s “cone of influence (COI),” the central segment of the wavelet in which changes in the underlying signal have the greatest impact on wavelet power (Torrence & Compo, 1998; see Figure

S5 for visualization of the COIs in this study). Qualitatively, the integration and modularity time courses might be consistent with a transient, while the shifts in strength and the number of modules appear longer-lasting and potentially indicative of more enduring changes to the network as a result of exploration. We return to these issues in the discussion.

3.5. The relationship between pupil-linked arousal and network integration and segregation

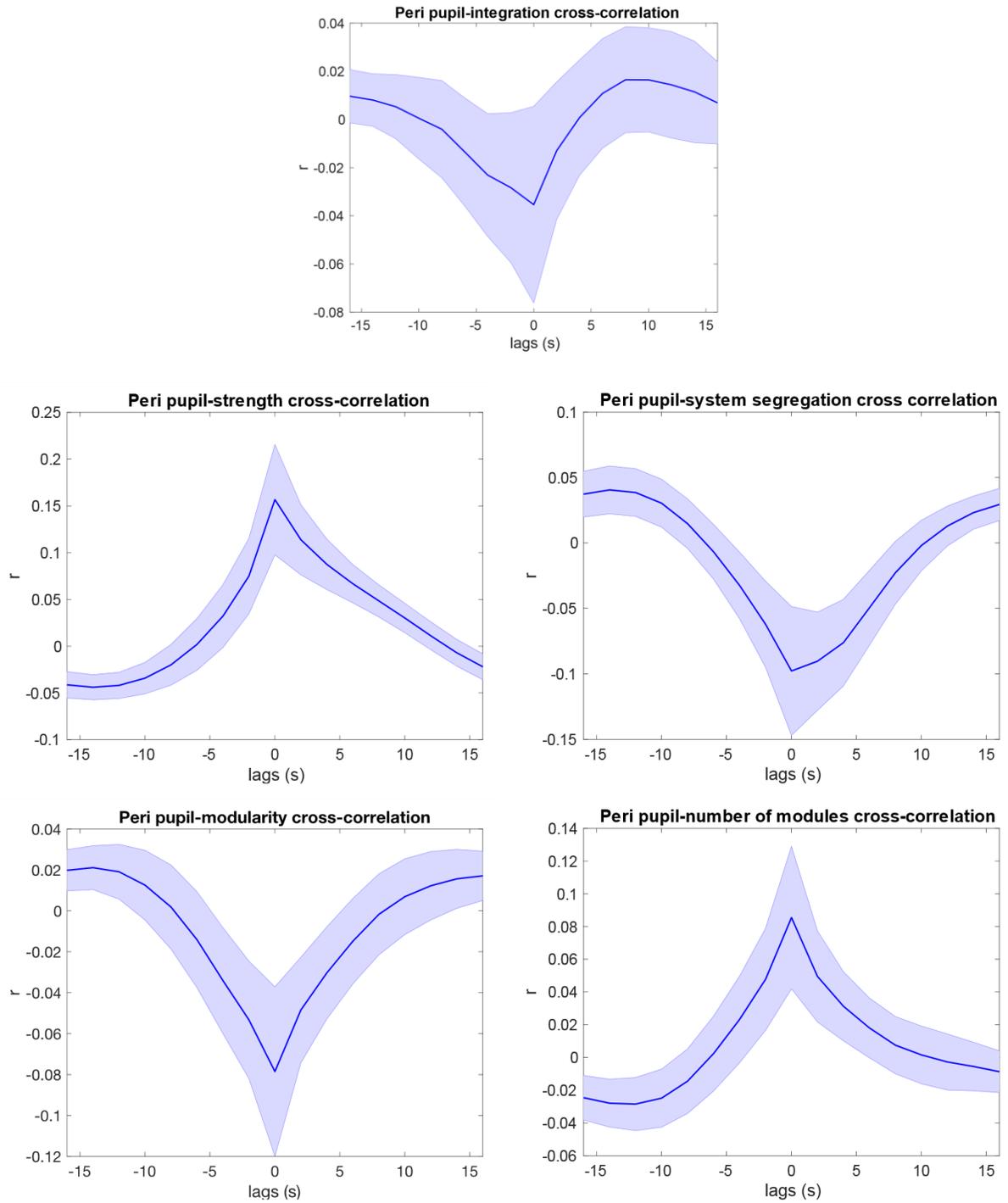


Figure 3.11. The cross-correlation between each network measure and the downsampled and low-pass-filtered pupil time course during the peri-explore period. Average cross-

correlations and SEMs were computed by first Fisher z -transforming the correlations at each lag, and then back-transforming for display.

Both pupil diameter and measures of network integration and segregation were modulated around exploration, raising the possibility that LC-NE activity influences integration during exploration, as hypothesized. To more formally assess this possibility, we computed the cross-correlation between pupil diameter and our network measures (see Methods). All measures demonstrated a peak at lag 0 (Figure 3.8), so we therefore assessed the significance of the zero-lag correlation across subjects. This relationship was weak overall, with the only significant correlation occurring for pupil–strength ($r_{ave} = .157$, $t(33) = 2.57$, $p = .015$). However, it was at trend for all other measures but integration (integration: $r_{ave} = -.035$, $t(33) = -0.88$, $p = .38$; system segregation: $r_{ave} = -.098$, $t(33) = -1.77$, $p = .085$; modularity: $r_{ave} = -.079$, $t(33) = -1.90$, $p = .066$; number of modules: $r_{ave} = .086$, $t(33) = 2.01$, $p = .052$). This finding replicates prior work demonstrating a positive association between pupil diameter and overall strength of functional connectivity at the block level (Eldar et al., 2013; Warren et al., 2016). Given that the other measures are all ultimately derived from connectivity strength, it may be that further noise introduced by those calculations—particularly those involving the computation of modularity, may have served to partially obscure these relationships. It may also be the case the effect of LC-NE activity during exploration is best characterized as influencing overall connectivity strength, though the other measures indicate that this effect is somewhat heterogenous. Yet taken together, these results suggest a role for LC-NE activity in the complex changes in network connectivity and topology around exploration.

3.6. Pupil diameter is not modulated by volatility condition

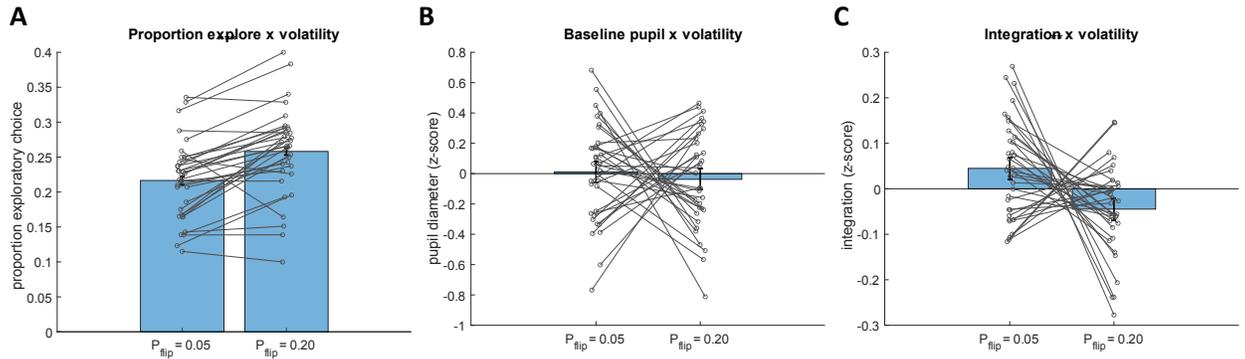


Figure 3.12. Effects of the block-level volatility manipulation. **A** proportion exploratory choice. **B** baseline pupil diameter. **C** integration. Integration was z-scored within subject for visualization, but analyses were performed on untransformed values. ** $p < .01$; *** $p < .001$.

We next assessed whether manipulating the volatility across blocks produced changes in subjects' exploratory behavior, pupil-linked arousal, and brain network integration. As predicted, subjects explored significantly more in high volatility blocks (Figure 3.9A; $\beta = 0.23$, $z = 4.74$, $p < .0001$). The magnitude of this effect, however, was smaller than expected, ($M_{diff} = 0.04$, or approximately 3 trials), despite the markedly differing rates of change per block. It may be that because the volatility changes were unsigned, subjects were not certain or aware enough of the block-level differences to strongly alter their behavior. Prior work also suggests that when subjects are not explicitly made aware of the full task structure, they are not always able to discover it (Payzan-LeNestour & Bossaerts, 2011).

Contrary to expectations, block-level baseline pupil diameter was not modulated by volatility condition (Figure 3.9B; $t(33) = 0.50$, $p = .62$). This was not a result of noise due

to gaze position or slow drifts in the pupil signal across blocks, as there was no difference when controlling for gaze and a 4th-degree polynomial over trials ($\beta = -0.01$, $t(33.12) = -0.09$, $p = .93$). Prior work has suggested using pupil dilation responses as a less noisy surrogate for tonic (baseline) pupil diameter, given the inverse relationship between tonic and phasic pupil dilation/LC responses (Eldar et al., 2013; Eldar, Niv, & Cohen, 2016). Indeed, these two measures were negatively associated in our sample ($\beta = -1.23$, $t(30.85) = -19.02$, $p < .0001$). However, dilation responses also showed no modulation by volatility, either alone ($t(33) = 1.30$, $p = .20$) or controlling for gaze position and slow drifts ($\beta = -0.02$, $t(33.01) = 1.03$, $p = .31$). Nor did dilation response predict the overall level of exploration across subjects ($r(32) = .157$, $p = .38$).

3.7. Volatility decreases block-level integration

Despite the absence of effects of volatility condition on arousal responses, volatility condition did significantly impact brain network integration in the predicted direction—integration was lower in high volatility blocks than low volatility blocks (Figure 3.9C; $t(33) = 2.82$, $p = 0.008$). We therefore also asked whether this effect was driven by particular cognitive systems or whether it was better characterized as a global phenomenon. The effect was qualitatively present across cognitive systems and was significant in all but the ventral attention system (all $ps < .05$), but only the dorsal attention and subcortical systems survived correction for multiple comparisons (Figure S6.; both $p_{\text{corrected}} = .045$). In contrast, none of the between-system interactions were significant at a corrected level (all $p_{\text{corrected}} > .18$). This suggests that the effect was widespread but somewhat heterogenous in the size of the effect among cognitive systems.

Because choice behavior varied to such a small degree at the group level between volatility conditions, it is unclear whether this difference could have been the driver of differences in integration; alternatively, the experienced volatility itself may have modulated brain network integration. Indeed, when entered together into the same model, the mean rate of exploration at each volatility condition was not a significant predictor of integration across subjects ($\beta = -0.002$, $t(64.75) = -0.10$, $p = .92$), while the mean rate of experienced changes was ($\beta = -0.02$, $t(46.49) = -2.17$, $p = .036$). This strongly suggests that the experience of a more changeable and uncertain environment drove the brain into a less integrated state.

To further rule out the possibility that exploratory choice and not volatility drove this difference, we revisited the peri-explore integration analysis, asking whether the volatility level may have blunted the response to exploration in the high volatility block, potentially contributing to reduced integration. Corroborating the block-level analysis, integration was significantly lower in the high volatility blocks on average during the peri-explore period ($\beta = -0.004$, $t(4550.65) = -2.49$, $p = .013$). However, there was not a significant difference in the peri-explore modulation between volatility conditions ($F(1.71, 4550.65) = 0.65$, $p = .57$), suggesting that this did not drive the block-level effect.

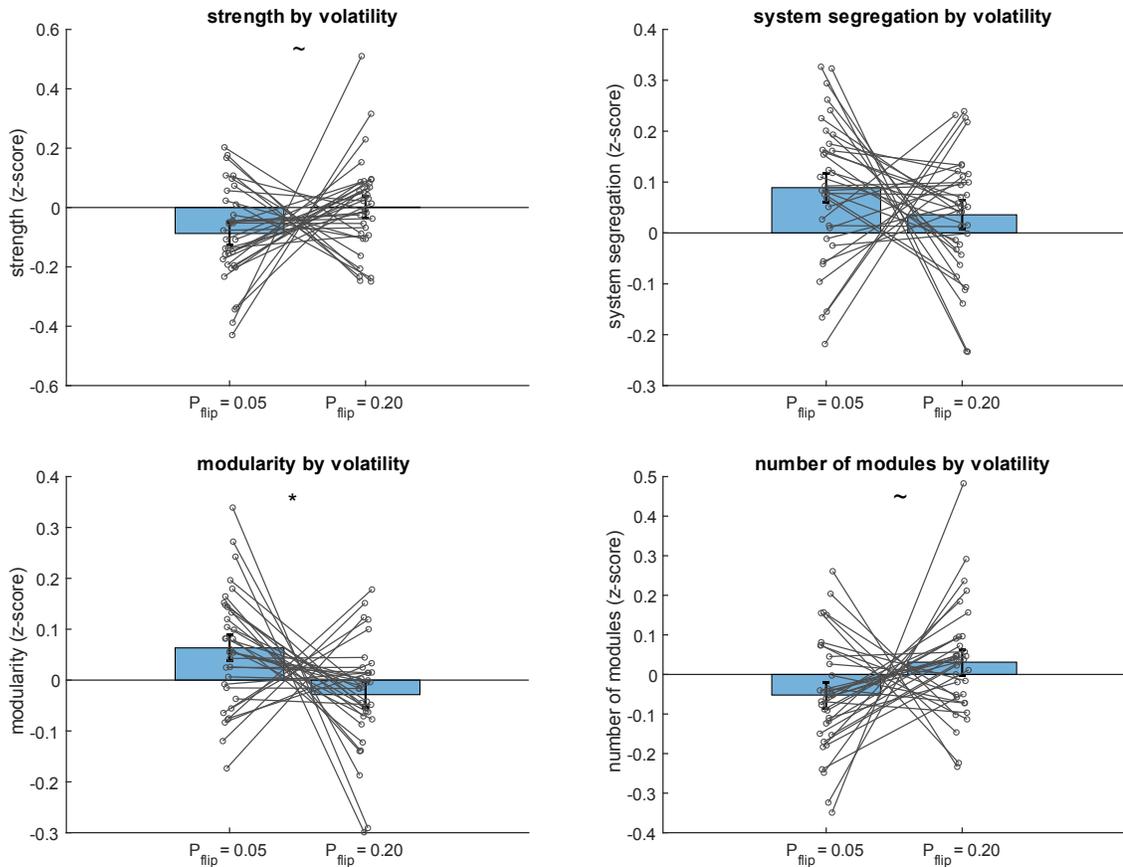


Figure 3.13. The effect of block-level volatility on average node strength, system segregation, modularity, and the number of modules. $\sim p < .10$; $* p < .05$.

Finally, to better characterize the volatility-driven change in integration, we examined changes in other measures of connectivity and topology (Figure 3.10). We found that overall node strength increased marginally ($t(33) = 1.91, p = .065$). System segregation was lower in the high volatility blocks, though this was significant relative only to the modules at each time point (Yeo cognitive systems: $p = .19$; modules: Figure S7, $t(33) = -2.23, p = .03$), indicating a slight shift toward more tightly connected communities. This was accompanied by a decrease in modularity ($t(33) = -2.63, p = .01$), as well as a marginally significant increase in the number of modules ($t(33) = 1.99, p = .055$).

Overall, then, increased volatility was associated with a slightly larger collection of modules that were more tightly integrated with respect to their connectivity but were more segregated relative to the cognitive systems present in each community. This result is in all respects a mirror image of the effect produced by exploration—albeit weaker—further emphasizing the distinct effects produced by volatility and exploration as well as the tight interrelationships among these measures in this task.

3.8. Assessing the relationship between block-level integration and arousal

Though pupil diameter was not modulated by volatility condition, it could still be the case that arousal levels moderated the effect of volatility on integration. This was not the case. There was no interaction between volatility condition and baseline pupil ($\beta = -0.001$, $t(54.93) = -0.21$, $p = .83$), nor was there a main effect of baseline pupil ($\beta = -0.001$, $t(33.68) = -0.32$, $p = .75$). Average baseline pupil diameter across the task was also not associated with average integration ($r(32) = -.093$, $p = .60$). Thus, in contrast to prior studies (Eldar et al., 2013; Warren et al., 2016), we do not find a relationship between pupil-linked arousal and network organization at the block level.

4. Discussion

Here we assessed the relationship between LC-NE-linked changes in pupil diameter, brain network integration, and behavior in the context of exploratory choice. Consonant with our predictions and corroborating previous findings (Jepma & Nieuwenhuis, 2011), we found that exploration induced a reliable increase in pupil diameter. This is in line with the adaptive gain theory of LC-NE function, which states that changes in tonic LC

firing mediate between states of exploration and exploitation (Aston-Jones & Cohen, 2005). We also examined changes in brain network integration around exploration. While our hypothesis that integration would be modulated around exploration was confirmed, the simple directionality of the hypothesis was not. Rather than finding strictly reduced integration, exploration-linked alterations in functional network architecture across a range of measures were consistent with a shift toward fewer, more weakly connected modules that were both more segregated in terms of connectivity and topology but also more integrated with respect to the diversity of cognitive systems represented in each module. Importantly, overall functional connectivity strength decreased, and changes in connectivity were associated with changes in pupil diameter, in line with the hypothesis that changes in LC-NE activity contribute to the dynamic reorganization of brain networks. These findings are the first to tightly link NE-associated arousal, brain network dynamics, and behavior in human subjects, going beyond prior studies, which relied on incidental variations in arousal or pharmacological manipulation assayed over longer periods of time. In so doing, this study has pushed the temporal grain at which sliding-window network analyses have been applied, indicating the possibility of using these methods to uncover finer-timescale changes when carefully coupled to behaviors of interest.

We also assayed whether block-level differences in environmental volatility would induce coupled changes in exploratory behavior, brain network integration, and pupil diameter. This manipulation was ultimately unsuccessful, as it elicited only weak differences in exploratory choice between volatility conditions and no differences in

pupil diameter. We did, however, find an unexpected association between functional brain network architecture and volatility condition—high volatility blocks were characterized by brain networks that were less integrated with respect to the diversity of cognitive systems present in each community but more integrated with respect to their connectivity and modularity. While we failed to confirm our predictions, this effect appeared separable from the effect of exploration and highlights the need to continue examining multiple contextual and neurobiological determinates of brain network dynamics, not just endogenous fluctuations during resting state (Medaglia, Lynall, & Bassett, 2015).

4.1. Complex peri-explore network dynamics

One factor that may be particularly important in driving the present results is the overall decrease in connectivity strength. Closely mirroring our findings, in a model of coupled oscillators, global decreases in coupling strength can lead to decreases in synchronization both within and between communities, as well as increases in modularity (Zhao, Zhou, Chen, Hu, & Wang, 2010). Changes in coupling strength have also been a target of modeling the effect of LC-NE activity on brain networks, which can lead to nonlinear changes in the degree of integration in the network (Shine et al., 2018a).

However, the complex changes in functional network architecture during the peri-explore period stand in contrast to some prior findings in the literature. For example, performing the cognitively demanding n-back task has been found to increase brain network integration as measured in the present study (Braun et al., 2015), as measured by the

diversity of intermodular connections (participation coefficient; Shine et al., 2016), and as measured by the average path length between nodes (global efficiency; Cohen & D’Esposito, 2016). It has also been found to decrease modularity (Cohen & D’Esposito, 2016; Vatansever et al., 2015) and system segregation (Cohen & D’Esposito, 2016)—both taken as measures of segregation—and decrease the number of modules (Vatansever et al., 2015). In the case of the n-back at least, all measures converge on a depiction of brain networks that have become more integrated (less segregated) in their connectivity and topology. Indeed, while integration and segregation can be measured separately (Deco, Tononi, Boly, & Kringelbach, 2015; Rubinov & Sporns, 2010), such measures display anticorrelations in both computational models (Deco et al., 2015) and empirical data (Cruzat et al., 2018), as is also implied by the findings from the n-back data across studies.

The divergence between these findings and the conflicting changes in integration and segregation found during exploration highlight the need to assess putative changes in integration across a range of tasks and measures. For example, a neural network model trained on multiple measures of segregation and integration was better able to predict performance across a range of tasks than the individual measures alone, suggesting that each contributes unique information (Bertolero, Yeo, Bassett, & D’Esposito, 2018). Moreover, as implied by our initial hypotheses, more integration—however defined—may not always be better. For example, performance in motor tasks has been shown to benefit from increased segregation of brain networks (Bassett, Yang, Wymbs, & Grafton, 2015; Cohen & D’Esposito, 2016). Indeed, it has been suggested that more modular brain

networks are of benefit in simple tasks that rely on segregation of processing and relatively isolated cognitive systems, while less modular networks are better in more complex tasks that require integrated processing (Yue et al., 2017).

All of this raises the question of what is the benefit of modulating integration in the context of exploration, which is not well-captured by the distinction between simplicity and complexity. Indeed, these changes in state occur in the context of the exact same task. Modeling suggests that networks constrained to be sparser and more modular in some cases are better at converging to the solution in a given task (Bernatskiy & Bongard, 2015) and better adapt to task changes (Clune, Mouret, & Lipson, 2013). Importantly, structural brain networks are not only modular, but also small-world, characterized by high clustering and short path lengths (Bassett & Bullmore, 2006). While small-world networks need not be modular, this property of the brain has been proposed to balance the segregated processing afforded by modularity with integrative processing afforded by more global connectivity (Bassett & Bullmore, 2006; Gallos, Makse, & Sigman, 2012). Interestingly, small-world topology has been shown to impact exploration and exploitation in the context of problem-solving networks. In such networks, agents attempt to find the best solution to a problem in parallel (e.g., guessing the number that yields the highest payoff), where individuals connected to each other in the network have access to one another's answers. Networks of human subjects as well as simulated agents display more exploration of the problem space in less connected networks due to greater segregation of information (Lazer & Friedman, 2007; Mason, Jones, & Goldstone, 2008). While fully connected networks excel in unimodal problem

spaces, small-world networks excel in multimodal problem spaces (Mason et al., 2008). Notably, some of the same benefits of structural connectivity can be obtained by changing the dynamics, such that agents can only occasionally view the solutions of their network neighbors (Bernstein, Shore, & Lazer, 2018; Lazer & Friedman, 2007). As may be expected, these results are highly dependent on the type of problem to be solved (Mason & Watts, 2012; Shore, Bernstein, & Lazer, 2015), and they come from networks at a far remove from brain networks. However, they suggest the intriguing possibility that dynamically increasing segregation in the brain during exploration may increase its ability to flexibly adapt when exploring new problem spaces or environments. On the other hand, the fact that the overall number of modules decreased, contributing to an increase in integration of different cognitive systems, may serve to balance this segregation by increasing the diversity of processing within each module. While these ideas are speculative by way of analogy to other networks, they suggest important areas for future research utilizing neural network models.

4.2. Specificity of network effects

While some studies have suggesting that LC-NE-linked modulation of network connectivity is relatively global, in keeping with the diffuse projections of LC (Eldar et al., 2013), others have uncovered heterogeneity in these effects and linked them to catecholamine receptor distributions (van den Brink et al., 2018, 2016b; Zerbi et al., 2019). Furthermore, recent work in rodents indicates that LC neuron projections and the interactions among LC ensembles are far more regionally specific with respect to their cortical targets than previously appreciated (Totah, Logothetis, & Eschenko, 2019).

We also found evidence for specificity—modulation of integration around exploration was most prominent in the default, dorsal attention, limbic, and frontoparietal systems and their interactions. While the default mode network was initially defined based on its decreased activity during task (Raichle, 2015), an increasing body of work suggests its relevance for task processing. In particular, it has been implicated in working memory (Vatansever et al., 2015), task switching (Crittenden et al., 2015), attentional shifting (Arsenault, Caspari, Vandenberghe, & Vanduffel, 2017), and creative cognition (Beaty, Benedek, Silvia, & Schacter, 2016). Of particular relevance to the present study, neurons in posterior cingulate—a key DMN node—have been implicated in performance monitoring (Heilbronner & Platt, 2013) and exploration (Pearson, Hayden, Raghavachari, & Platt, 2009). There is also prior evidence of dynamic interactions between default, frontoparietal, and dorsal attention systems, with the frontoparietal network potentially regulating activity in the other two networks in order to adjust the balance between internally-generated (default) and externally-directed (dorsal attention) processing (Beaty et al., 2016; Dixon et al., 2017, 2018; Smallwood, Brown, Baird, & Schooler, 2012). Furthermore, interactions among the limbic, attentional, and LC-NE systems appear to modulate attention, learning, and memory for salient or motivationally relevant events (Clewett & Murty, 2019; Gallagher & Holland, 1994; Mohanty, Gitelman, Small, & Mesulam, 2008). The Leapfrog task itself has been associated with both prefrontal function and arousal (Blanco et al., 2015; Otto et al., 2014). While we can only speculate about the role of these networks and their interactions in the present study, they may reflect the coordination of monitoring, decision-making, and attentional processes in service of flexibly shifting between exploitation and exploration based on ongoing

estimates of the relative value of exploring. Regardless, the specificity of these effects provides further motivation for examining the role of LC-NE activity in modulating brain network connectivity within specific contexts.

4.3. Pupillary response to exploratory state

While it was not a primary goal of the study, our results also bear strongly on the role of LC-NE-linked arousal in mediating between exploration and exploitation. Despite the long-standing hypothesis that tonic LC activity mediates between these states (Aston-Jones & Cohen, 2005), relatively few studies have examined this relationship, though most have found support for such a relationship (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Hayes & Petrov, 2016; Jepma & Nieuwenhuis, 2011; Kane et al., 2017; cf. Jepma, Te Beek, Wagenmakers, van Gerven, & Nieuwenhuis, 2010; Warren et al., 2017). Pupil diameter is sensitive to several non-luminance-mediated factors, including uncertainty and surprise (Alamia et al., 2019; Friedman et al., 1973; Jepma & Nieuwenhuis, 2011; Lavín et al., 2014; Nassar et al., 2012; Preuschoff et al., 2011; Qiyuan, Richer, Wagoner, & Beatty, 1985; Urai, Braun, & Donner, 2017), as well as mental load or task difficulty (Alnaes et al., 2014; Hess & Polt, 1964; Kahneman & Beatty, 1966; Wahn et al., 2016). Notably, past task designs used to test the relationship between LC-NE activity and exploratory state do not clearly differentiate states of exploration from these other factors. For example, in drifting bandits, exploratory choice tracks with the entropy of the option values (Jepma & Nieuwenhuis, 2011), and a canonical study of exploration, operationalized as task disengagement, utilized increases in task difficulty to promote disengagement (Gilzenrat et al., 2010). It could thus be the

case that pupil diameter in these studies is more related to other variables than to exploration per se—indeed, in both cases pupil diameter was argued to closely track expected utility, a putative signal of when to initiate exploration (Aston-Jones & Cohen, 2005). While it is an empirical question whether states of exploration reduce to states of uncertainty or low utility, the information gained by exploration has utility in and of itself, despite the opportunity costs associated with potentially lower payoffs (e.g., directed exploration; Gershman, 2018; Kaelbling, Littman, & Moore, 1996; Knox et al., 2012; Wilson, Geana, White, Ludvig, & Cohen, 2014). Furthermore, mice demonstrate elevated pupil diameter during exploratory behaviors that are not associated with immediate payoffs (McGinley et al., 2015). Exploratory states would thus seem to be at least somewhat separable from these other factors, and potentially heterogenous in nature.

The simplified nature of the Leapfrog task mitigates these concerns; the option values change in a highly constrained way, meaning the only uncertainty/difficulty lies in the decision of when to explore, given the rate of change in the environment (Knox et al., 2012). Crucially, we found no anticipatory increase in pupil diameter on the trials leading up to the explore trial. Instead, pupil diameter appeared to be elevated in response to the decision to explore itself. The canonical pupillary response function has an approximately one second lag to peak and returns to baseline after about two seconds (Hoeks & Levelt, 1993). Choice on exploit trials closely followed this pattern (Figure 3.2), suggestive of phasic LC-NE activity, while choice on explore trials remained elevated for several seconds following the explore choice, suggestive of a tonic (though brief) elevation in

LC-NE activity. Importantly, response to a change in outcome did not drive this effect, ruling out a role for surprise, and the explore response was not sensitive to volatility condition, suggesting it also was not due to greater uncertainty in the outcome of explore choices. This conclusion must be qualified, however, by the relatively weak sensitivity of our subjects to the volatility manipulation. Given that the pupillary response has been shown to be modulated by probabilities and at least qualitatively demonstrates more extended responses to low probability events (Alamia et al., 2019; Qiyuan et al., 1985), we cannot completely rule out this possibility, but we tentatively propose that the pupillary response to exploration in this case reflects the shift into an exploratory state itself, apart from decision variables contributing to the decision to explore. Furthermore, this suggests that increased arousal in this case was a consequence of the decision to explore, rather than its cause. In keeping with this, we failed to replicate a prior finding that average tonic pupil diameter is associated with rates of exploration across subjects in a drifting bandit task (Jepma & Nieuwenhuis, 2011). It may be the case that individual differences in tonic arousal are more associated with random exploration of the sort elicited in drifting bandits (Daw et al., 2006; Wilson et al., 2014), but that directed exploratory decisions of the sort elicited by the Leapfrog task can lead to intentional shifts into high-arousal exploratory states.

4.4. Limitations and future directions

While this study identified exploration-induced modulation of brain network connectivity on a fairly fine temporal scale, there are a number of caveats that bear consideration. First, the low-frequency nature of the continuous wavelet coherence analysis makes it

difficult to infer the exact nature of the underlying neural activity. Indeed, filtering, including the use of wavelets, can distort the timing of the underlying signals (de Cheveigné & Nelken, 2019; Yael, Vecht, & Bar-Gad, 2018). This is particularly evident when examining the post-explore pupil time course after low-pass filtering, which displays a different character than prior to filtering (Figures 3.4B, 3.5B). Thus, while our analyses provide evidence of an exploration and LC-NE-linked modulation, the exact nature of the modulation—its timing and directionality—may be quite different than that uncovered here. On the other hand, wavelet analysis has benefits over correlation-based methods in robustness to noise and temporal autocorrelation (Zhang, Telesford, Giusti, Lim, & Bassett, 2016).

Relatedly, we took substantial steps to address temporal autocorrelation in our analyses, including the use of GAMMs, AR1 error models, and corrected correlation Z-scores. Although the impact of temporal autocorrelation—particularly in nonstationary time series—has long been recognized in fields such as economics and statistics (Granger & Newbold, 1974; Johansen, 2012; Phillips, 1986; Yule, 1926), and univariate analyses of fMRI data correct for non-independence in the residuals of fMRI GLM analyses due to autocorrelation (Monti, 2011), autocorrelation has not always been taken into account in psychological and neuroscientific analyses, including in analyses of pupil–network relationships. This potentially threatens not only statistical inference (i.e., inflated Type I error rate), but also in some cases the validity of the parameter estimates themselves (i.e., spurious correlation). That said, there has been disagreement as to the severity of the autocorrelation problem, likely owing to differences in the underlying signals, the length

of the time series, and the assumptions made about the autoregressive processes (Afyouni et al., 2019; Arbabshirani et al., 2014; Baayen et al., 2017; Dean & Dunsmuir, 2016; Elber-Dorozko & Loewenstein, 2018; Honari, Choe, Pekar, & Lindquist, 2019; Leonardi & Van De Ville, 2015). We have chosen to take this problem seriously, though other solutions, such as prewhitening or the use of ARIMA models could have been used, as is recommended by some of these authors. We did not use these methods here because we did not want to eliminate low-frequency signal components (Afyouni et al., 2019; Pyper & Peterman, 1998), but future work should assess the impact of various mitigation strategies not only on functional connectivity itself, but its relation to other signals of interest such as pupil diameter. It may also be worth investigating the use of clustering (Khambhati, Mattar, Wymbs, Grafton, & Bassett, 2018; Liu, Zhang, Chang, & Duyn, 2018; Medaglia et al., 2018) or deconvolution (Karahanoğlu, Caballero-Gaudes, Lazeyras, & Van De Ville, 2013; Wierda, van Rijn, Taatgen, & Martens, 2012) techniques to aid in addressing both issues of temporal precision and autocorrelation.

While we have attributed the peri-explore modulation to a result of exploration under the putative influence of NE, both of these assumptions must be examined in more detail in future studies. Given our task design and limits on the amount of explore trials per subject, we cannot completely disentangle effects of exploration from effects of change, uncertainty, and overall volatility, though we made several attempts to do so.

Furthermore, in the Leapfrog paradigm bouts of exploratory choice are usually on the order of a single trial. Designs that provoke more extended exploratory states may help to overcome issues related to temporally isolating the effects of exploration. Additionally,

we cannot separate effects of exploration from more general effects of attentional shifting. While LC-NE-linked effects on attentional processes are well-known and in some sense are partly constitutive of its influence on exploratory state (Aston-Jones & Cohen, 2005; Corbetta, Patel, & Shulman, 2008; McGinley et al., 2015; Sara & Bouret, 2012), exploration has been isolated from switching at the single-neuron level (Pearson et al., 2009), so it will be important to better delineate the boundaries of these different processes/states in the future.

Other neuromodulators such as dopamine and acetylcholine have also been implicated in coordinating brain network dynamics (Birn et al., 2019; Roffman et al., 2016; Shafiei et al., 2019; Turchi et al., 2018; Záborszky et al., 2018) and have been implicated in uncertainty and exploration (Beeler et al., 2010; Fiorillo, Tobler, & Schultz, 2003; Yu & Dayan, 2005). Acetylcholine in particular also influences pupil diameter (Reimer et al., 2016), meaning that we cannot rule out its impact in the present results, as is the case in all studies utilizing pupil diameter as a proxy for LC-NE activity. Finally, other mechanisms, such as thalamic regulation, have been linked to the control of cortical connectivity (Halassa & Kastner, 2017). Given that we could not link the network effects of volatility to pupil diameter, this highlights the need look beyond neuromodulators for other mechanisms of brain network reconfiguration.

In sum, we have demonstrated a relationship between exploration, pupil-linked arousal, and brain network dynamics. We argue that forming linkages between functional connectivity, behavior, and physiological markers such as pupil diameter represents a

promising path forward for understanding the effects of neuromodulatory actions on brain network dynamics and their impact on cognitive processing.

IV: GENERAL DISCUSSION

This thesis presented two case studies of the neurobiological substrates of endogenous flexibility, the ability to adapt behavior without explicit cues to do so. In particular, we focused on the neural substrates that influence the balance between stability and flexibility, including the roles of the prefrontal cortex and the neuromodulators dopamine and norepinephrine. The results of these studies contribute to an expanding literature that is illuminating the interconnectedness of learning and control processes in supporting adaptive behavior. Control is implemented most effectively when deployed based on learned estimates of environmental variables (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Collins & Koechlin, 2012; Jiang et al., 2015; Jiang, Heller, & Egner, 2014; Shenhav et al., 2013; Yu & Cohen, 2009). Similarly, learning is most effective when it integrates top-down control signals that serve to stabilize behavior in line with current goals with bottom-up information that may contain signals that the environment has changed, potentially necessitating changes in the current control state, current goals, and/or current estimates of environmental variables (Cohen et al., 2007; Daw et al., 2005; Nassar et al., 2010; Pearson et al., 2011; Yu & Dayan, 2005). When there is a mismatch between the environment or task at hand and the current control state, this can produce behavior that is either too stable or too flexible, hindering performance. Importantly, the balance between stability and flexibility is powerfully influenced by the actions of neuromodulators such as dopamine and norepinephrine (Aston-Jones & Cohen, 2005; Cools & D'Esposito, 2009; Nassar et al., 2012; Yu & Dayan, 2005).

In Chapter 2, we examined the role of frontostriatal circuits—and their modulation by DA—in instructed reinforcement learning. When instructions are inaccurate, a conflict is generated between the top-down information provided by the instructions and bottom-up reward information provided through experience. This conflict provides a window on the balance between stability and flexibility via the extent to which instructions bias learning. Consistent with increased top-down regulation of reinforcement learning, subjects who received anodal stimulation over PFC demonstrated greater bias relative to sham, though this effect was present only early in training. This provides the first causal evidence of a role for PFC in instructed RL, indicating that too much PFC-mediated stability in following the instructions is detrimental to learning. We also replicated the finding that the COMT Met allele is associated with increased instructional bias (Doll et al., 2011) and further demonstrated that variation in DAT1 has similar effects to variation in COMT, with 9-repeat carriers demonstrating increased bias relative to 10-repeat homozygotes. Intriguingly, COMT Met homozygotes who were also DAT 9-repeat carriers demonstrated markedly higher bias than all other genotypic groups. These results support the idea that the balance between PFC and striatal DA, rather than the functioning of the PFC alone, determines the balance between stability and flexibility (Cools & D'Esposito, 2009). Finally, we fit computational models to subjects' data to better characterize the mechanisms underlying instruction bias. A novel choice bias model, in which instructions influence decision-making rather than learning, was found to best account for subjects' behavior. Together, these data add to the growing literature documenting both costs and benefits of cognitive control (Chrysikou et al., 2014), and

indicate that neurobiological differences in the stability-flexibility balance can lead to mismatches with the task at hand, in this case producing behavior that is overly stable.

Chapter 3 viewed the balance between stability and flexibility through a different behavioral and neuromodulatory lens, assaying the relationship between choice behavior, brain network dynamics, and LC-NE activity in the context of exploration and exploitation. The balance between exploration and exploitation is perhaps the paradigmatic example of the tradeoff between stability and flexibility (Aston-Jones & Cohen, 2005; Mehlhorn et al., 2015). In the case of the Leapfrog task, stably exploiting for too long means likely missing out on higher payoffs from the unchosen option, while exploring too often entails missed opportunities to exploit the better option. Consistent with the view that higher tonic LC-NE activity promotes exploration (Aston-Jones & Cohen, 2005) and in line with prior work (Jepma & Nieuwenhuis, 2011), subjects demonstrated increased pupil diameter after exploratory choices. We also found modulations of brain network dynamics around exploration across several measures. These changes were associated—albeit weakly—with changes in pupil diameter, with the most reliable effect occurring for overall connectivity strength. These results tentatively support our hypothesis that modulation of brain network integration by LC-NE activity is a mechanism by which NE shuttles the brain between states of exploration and exploitation. However, while we predicted that integration would strictly decrease with exploration, we instead found a complicated pattern of effects; around exploration, network measures indicated a move toward lower overall connectivity and fewer, more weakly connected modules that were both more segregated when measuring connectivity

and topology but more integrated when measuring the diversity of cognitive systems within each module. These results perhaps augment, rather than clarify, an already conflicting literature on the role of LC-NE-linked modulations of functional connectivity (e.g., Shine, van den Brink, Hernaus, Nieuwenhuis, & Poldrack, 2018) and suggest important considerations for future work, to which we now turn.

1.1. Future directions

There is still much to be discovered with regard to the neural mechanisms of stability and flexibility. Within instructed reinforcement learning, it remains an open question how best to characterize the nature of the instructional bias. While we found evidence for a choice bias rather than a learning bias, as reviewed in Chapter 2 there is considerable variability across studies in whether the bias is attributed to learning or choice. One approach to this question that should be taken is to ask how recoverable and separable learning and choice bias parameters are in simulated data. It may be the case that current paradigms and models cannot actually adjudicate between these possibilities, which would necessitate new approaches. For example, a learning bias might suggest that memory systems other than striatally-mediated RL would be affected, such as episodic encoding of reward information, which has been recently demonstrated to influence choice (Bornstein, Khaw, Shohamy, & Daw, 2017). Therefore, paradigms that triangulate the bias by manipulating or assessing different aspects of learning and choice (e.g., memory for episodic information related to the instructed stimulus) might prove more fruitful than modeling alone.

More generally, given the putative role of prefrontally-mediated working memory processes in implementing the bias, it is of interest to ask whether individual differences, such as those dictated by COMT genotype, are stable across tasks. In particular, while the COMT Met allele has been negatively associated with flexibility in instructed RL and reversal learning (Doll et al., 2011; Krugel et al., 2009), it has been positively associated with other types of flexibility, including the use of model-based reinforcement learning and directed exploration (Doll et al., 2016; Frank, Doll, Oas-Terpstra, & Moreno, 2009). That is, stability and flexibility themselves are multiply determined. Directly comparing performance across tasks may thus help identify the specific mechanisms that benefit (hinder) performance in specific contexts. Stronger or more developed cognitive control is thought to benefit tasks requiring more filtering, higher levels of abstraction, and/or more proactive use of control (Chrysikou et al., 2014; Munakata et al., 2012). One possibility is therefore that types of flexibility that are more driven by bottom-up input, such as discarding false instructions or completing reversals, will benefit from less top-down control, while those that require planning, maintaining and computing over many task variables, or inhibiting prepotent responding, will benefit from more top-down control. Finally, the performance of COMT Met/Met:DAT 9-repeat carriers suggests that certain subpopulations may be much more rigid in their stability or flexibility than others. Further study of such individuals in larger cohorts might thus have implications for the study of psychiatric conditions that feature inflexible thoughts or behavior, such as depression and obsessive-compulsive disorder (Gruner & Pittenger, 2017; Levin et al., 2013; Remijnse et al., 2013).

Regarding the role of LC-NE activity in adaptively modulating brain network connectivity, the disparate results across studies suggest an urgent need for additional computational modeling and matched experimental investigations. While relatively simple neural network models of the effect of LC-NE-associated gain modulation have had success in predicting behavioral performance (Eldar et al., 2013, 2016), predictions regarding connectivity changes have been inconsistent across studies, perhaps owing to very different modeling paradigms (Eldar et al., 2013; Shine et al., 2018a). If connectivity is to provide any insight into cognitive function, models that combine some level of biological plausibility with some level of topological similarity to the human brain, in the context of specific tasks, will likely be necessary. Additionally, functional connectivity and its derivatives are relatively nonspecific measure for indexing changes in cognitive processing. For example, it is difficult to know what to conclude from a gain-mediated increase in functional integration in a task-free model of coupled oscillators (Shine et al., 2018a). One possible future direction would be to utilize measures such as informational connectivity (Anzellotti & Coutanche, 2018; Coutanche & Thompson-Schill, 2013), which could be used to ask how changes in task-associated multivoxel patterns covary in simulated and actual brain networks in concert with changes in LC-NE activity and measured behaviors. In the case of exploration, this requires assaying not only exploratory choice, but using paradigms that allow for the assessment of the consequences of exploratory states on information processing and behavior. For example, exploration has been shown to induce a greater reliance on bottom-up stimulus salience in macaques (Ebitz & Moore, 2016). Finally, modeling might be a good place to gain some insight into the effect of pharmacological manipulations, which have produced

some results that conflict with inferences made based on pupil diameter (Jepma et al., 2010; Warren et al., 2017b), potentially due in the case of Atomoxetine to influences on both tonic and phasic NE (Bari & Aston-Jones, 2013).

Relatedly, it will also be critical to make progress in better understanding the inverted-U-shaped effects of NE (Arnsten, 2011; Aston-Jones & Cohen, 2005; McGinley et al., 2015), which may be another reason for disparate results across studies. For example, studies of attention suggest that optimal performance is attained at moderate levels of pupil-linked arousal (e.g., van den Brink et al., 2016a). Therefore, in any given study one might find a negative relationship, a positive relationship, or no relationship between performance and arousal, depending on the distribution of where study participants fall on the inverted-U. One possibility will be to use converging measures, such as salivary alpha amylase, to better characterize basal levels of NE (Warren, van den Brink, Nieuwenhuis, & Bosch, 2017a). Characterizing both sides of the U is also critical for understanding the balance between stability and flexibility. For example, high LC-NE activity has been proposed to focus processing on salient features, while low activity has been proposed to facilitate more integrative processing (Eldar et al., 2016), in much the same way that variations in cognitive control have been proposed to adjust the level of top-down filtering of information (Chrysikou et al., 2014; Shimamura, 2000). Mirroring the discussion of the role of prefrontal function and its modulation by DA, it is thus worth asking what types of flexibility and stability benefit from being located at different points on the U. Finally, like DA, NE also has inverted-U shaped effects on prefrontal function (Arnsten, 2011). Therefore, one final question is whether variation in prefrontal DA

levels interact with variation in NE-linked arousal to produce different optima in different subjects.

1.2. Conclusion

In *The Time Machine*, H.G Wells (1895) wrote:

“It is a law of nature we overlook, that intellectual versatility is the compensation for change, danger, and trouble. An animal perfectly in harmony with its environment is a perfect mechanism. Nature never appeals to intelligence until habit and instinct are useless. There is no intelligence where there is no change and no need of change. Only those animals partake of intelligence that have to meet a huge variety of needs and dangers” (ch. XIII, para. 3).

While intelligence may be debated, the need to balance stability and flexibility is undeniable. Even the simple nematode *C. elegans* must adjudicate between exploration and exploitation, and intriguingly this process is influenced by catecholamines (Bendesky, Tsunozaki, Rockman, Kruglyak, & Bargmann, 2011). The role of catecholamines in behavioral and neural flexibility may thus be nearly as evolutionarily ancient as animals' need for flexibility itself. It is perhaps one of the greatest clichés of our time that the need for adaptability is greater than ever in response to ever-increasing societal and technological change. Understanding the neurobiological and computational bases of stability and flexibility may thus provide a window into the past as well as a blueprint for optimizing flexibility in the future.

1. Supplementary Methods

1.1 Task Procedure

Subjects completed an instructed probabilistic selection task (iPST), presented on a 13” laptop computer via PsychoPy (Peirce, 2009). This task required subjects to learn the value of symbols initially presented in 3 pairs (AB, CD, EF; see Table 1, main text). Within each pair, one symbol had a higher probability of reward, and subjects were expected to learn to select the more highly rewarded symbols via feedback learning. Symbols were rendered as Japanese Hiragana characters and the assignment of Japanese character to underlying stimulus was randomized across subjects.

While seated in front of the computer, subjects first read the following instructions:

Thank you for participating! Two black symbols will appear simultaneously on the computer screen. One symbol will be “correct” and the other will be “incorrect”, but at first you won’t know which is which. There is no ABSOLUTE right answer, but some symbols have a higher chance of being correct than others. Try to pick the symbol that you find to have the highest chance of being correct. You’ll have to figure out which symbols to select by testing them out. Note: the side of the screen on which a symbol appears does not affect its chances of being correct. Now you will be introduced to the symbols.

Each symbol was then presented individually for 5 sec each. When symbol D was presented the screen also displayed the following false advice: “This symbol has the best chance of being correct.”

Subjects were then tested on how many stimuli appear on every trial and how to choose the stimulus on the left or the right. Each symbol was then presented again for 5 sec and subjects were directed to press a key when the instructed symbol D was presented. The instructions restarted from the beginning until all questions were answered correctly.

During the training phase, subjects had to learn the value of each symbol via probabilistic feedback. On a given trial, subjects saw one of the three symbol pairs, side counterbalanced. Trials began with a fixation cross, followed by the stimulus display. Once a response was made, the selected symbol was highlighted via a square border, colored green for positive feedback and red for negative feedback. Additionally, symbolic feedback in the form of a green checkmark for positive feedback and a red cross-out mark for negative feedback was displayed centrally below the two symbols. Feedback was only provided for the selected option. In order to ensure consistency across subjects in the duration of the task relative to stimulation, all trials were fixed in length and proceeded as follows: 300 ms fixation, 2000 ms response window, 200 ms highlight time, 900 ms feedback time. This was followed by a variable ITI (minimum 800 ms) calculated to bring the total duration of trial + ITI to 4200 ms. If subjects failed to make a response during the response window, a blue question mark was displayed in lieu of feedback for the remainder of the trial.

Subjects completed 4 training blocks. Each block contained 20 repetitions of each pair, for a total of 60 trials per block and 240 total training trials. Trial order and feedback were randomized within each block. Feedback was randomized such that within a block, each symbol was assigned reward at a rate equal to its underlying probability of reward (i.e., if the subject always chose symbol A, it would result in positive feedback on 16 of the 20 trials, for a $p(\text{reward}) = 0.8$). Feedback was also assigned in a complementary fashion within symbol pairs, such that in trials on which one symbol was assigned positive feedback the other symbol in the pair was assigned negative feedback.

After completing the training phase, the test phase began with the following instructions:

It's time to test what you've learned! During this set of trials you will NOT receive feedback (correct or incorrect) on your responses. If you see new combinations of symbols in the test, please choose the symbol that "feels" more correct based on what you learned during the training sessions. If you're not sure which one to pick, just go with your gut instinct!

During the test phase, all possible symbol pairings were presented (e.g., AB, AC, AD, AE, AF, ...) without feedback. Each pair was presented 6 times, for a total of 90 trials. Order was randomized across subjects.

1.2. Genotyping

DNA samples were collected via Oragene saliva kits (DNA Genotek) and extracted using the Chemagen MSMI DNA Extraction system. For the COMT Val158Met SNP, Taqman 5' nuclease PCR primers and probes were utilized (Life Technologies). Each probe

consisted of an oligonucleotide with a fluorescent reporter dye, a non-fluorescent quencher and minor groove binder (MGB). Allele-specific cleavage of probes was detected using different reporter dyes for each probe (6FAM and VIC fluorophores for each allele), with separate wavelength maxima. PCR amplifications were set up in a 384-well plate format in total volume of 5 μ l, containing 2.5 μ l 2X universal master mix, 0.25 μ l 1X primer and probe from ABI and 2.25 μ l of DNA at a concentration of 5 ng/ μ l. Water as a negative control was included in each 384-well plate. PCR was performed in QuantStudio 12K Flex Real-Time PCR System (ABI). After an enzyme activation step for 10 min at 95 °C, 60 two-step cycles were performed; 15 sec denaturation at 95 °C followed by 1 min annealing/extension at 60 °C for all variants. After PCR, end-point fluorescence levels of 6FAM and VIC were measured automatically in each well using V1.2.2 manufacturer's custom software (ABI). Allelic discrimination results were then graphed on a scatter plot contrasting reporter dye fluorescence (i.e., Allele X vs. Allele Y).

For the DAT1/SLC6A3 VNTR, extracted DNA was amplified using DAT1 VNTR specific primers (Forward primer: 5'-6FAM-TGT-GGT-GTA-GGG-AAC-GGC-CTG-AG-3'; Reverse primer: 5'-CTT-CCT-GGA-GGT-CAC-GGC-TCA-AGG-3'; ABI #450007) utilizing the Roche Expand High Fidelity PCR System (#04738268001).

Capillary electrophoresis was performed on the ABI 3130xl DNA Analyzer running POP7 polymer. One μ l of amplified sample was suspended in 9 μ l of Hi-Di Formamide (ABI #4311320) and 0.5 μ l of Genescan-600 LIZ Size Standard v2.0 (ABI #4408399) and denatured at 95 °C for 2 min then placed on ice for an additional 2 min before

loading onto the instrument. After electrophoresis, samples were analyzed using ABI Genemapper 4.0 software (Life Technologies).

2. Supplementary Results

Here we report the results of between-group parameter comparisons of the decision bias model (see sections Methods: Computational Modeling and Results: Computational Modeling of the main text for modeling details). These analyses complement the behavioral analyses in the main results of the paper, asking whether genotype and stimulation groups differ in the degree of their choice bias, as quantified by the ρ parameter of the model. See Supplementary Tables 1 and 2 for average parameter estimates for each group.

2.1. COMT

Mirroring the training phase behavioral results, we found a significant effect of COMT status on ρ ($F(2,99) = 3.31, p = .04$). The Met/Met group had a significantly larger bias than both Val/Val ($t(99) = 2.45, p_{corrected} = .049$) and Val/Met ($t(99) = 2.34, p_{corrected} = .049$). There was no difference between Val/Val and Val/Met ($t(99) = 0.34, p_{corrected} = .74$). We also found a significant gene-dose effect, whereby increasing Met alleles lead to increases in ρ ($r(100) = .21, p = .04$).

Test phase fit results were similar to those on the Avoid-D/Avoid-F measure. There was not a significant effect of COMT on ρ ($F(2,99) = 2.10, p = .13$). There was, however, a significant gene-dose effect ($r(100) = .20, p = .04$), whereby increasing Met alleles were associated with increasing bias.

2.2. DAT

Though 9-repeat carriers were on average fit with a higher value of ρ than 10/10 homozygotes ($M_{9c} = 0.20$, $M_{10/10} = 0.12$), this difference was only significant at a trend level ($t(100) = 1.90$, $p = .06$). Additionally, no other model parameters better explained the difference in instructed training phase performance (all $ps > .45$).

The inability to find a significant difference in model parameters despite a significant difference in behavioral performance could indicate that the decision bias model merely fails to capture the relevant difference between DAT groups. It may be the case, however, that noise in the parameter estimates masks a significant group difference. One common source of noise in such estimates is correlations among the parameters. We therefore asked whether 9-repeat carriers would have a significantly greater value of ρ , controlling for the other model parameters. Indeed they did ($\beta = 0.10$, $t(97) = 2.68$, $p = .009$)².

We did not find that DAT modulated the ρ parameter at test ($t(100) = 1.03$, $p = .31$).

However, examining the other parameters, we found that 9-repeat carriers had a significantly lower learning rate for losses (α_l) than 10/10 homozygotes ($t(100) = -2.75$, $p = .007$). No other differences were significant (all $ps > .14$). This difference in α_l is in keeping with the main effect of DAT on Avoid-D/Avoid-F performance in the absence of

² In light of this finding, we repeated all other group comparisons of the ρ parameter, controlling for the other parameters. The significance of all other comparisons were largely unchanged, excepting following: the COMT gene-dose effect at test fell to a trend level ($p = .055$), and the effect of DAC at test rose to a trend level ($p = .052$).

a significant interaction, though only the difference in Avoid-D was individually significant.

One question raised here is why a lower learning rate for losses would produce worse performance, when past investigations have demonstrated that a lower learning rate for losses can produce better performance in avoidance learning due to more stably learned values (Frank et al., 2007). One possibility is that the 9-repeat carriers were to a greater extent fit by *very* low α_1 , such that learning was impaired. Indeed, a greater proportion of 9-repeat carriers were fit with $\alpha_1 < .01$ compared to 10/10 homozygotes (9c: 45.7%, 10/10: 25.4%; $p = .046$, Fisher's Exact Test).

If low α_1 impaired learning, this should be reflected in the Q-values produced by the model. This was the case. In addition to a main effect of Symbol, indicating that overall Q-values were differentiated among the symbols ($F(5,500) = 95.49, p < .0001$), we also found a main effect of DAT ($F(1,100) = 9.41, p = .003$), qualified by a Symbol x DAT interaction ($F(5,500) = 2.83, p = .016$). Post-hoc tests revealed that all symbol values were inflated in the 9-carrier group relative to 10/10, and all these differences were significant except that for symbol A ($A_{9c-10/10}: M = 0.06, t(171.82) = 1.06, p_{corrected} = .29$; $B_{9c-10/10}: M = 0.17, t(171.82) = 2.89, p_{corrected} = .017$; $C_{9c-10/10}: M = 0.14, t(171.82) = 2.36, p_{corrected} = .04$; $D_{9c-10/10}: M = 0.19, t(171.82) = 3.34, p_{corrected} = .005$; $E_{9c-10/10}: M = 0.16, t(171.82) = 2.69, p_{corrected} = .02$; $F_{9c-10/10}: M = 0.21, t(171.82) = 3.65, p_{corrected} = .002$).

These results suggest that the distortion of Q-values in the 9-carrier group affected negatively valued stimuli (B, D, F) greater than positively valued stimuli (A, C, E). This

in turn could have affected the overall spread in value between negative and positive stimuli. Indeed, while both groups valued positive stimuli more than negative stimuli (9c: $M_{\text{pos-neg}} = 0.21$, $t(100) = 7.03$, $p_{\text{corrected}} < .0001$; 10/10: $M_{\text{pos-neg}} = 0.28$, $t(100) = 13.12$, $p_{\text{corrected}} < .0001$), this difference was reduced in 9-repeat carriers ($F(1,100) = 3.96$, $p = .049$). This still does not explain, however, why 9-carriers were (nonsignificantly) more impaired on Avoid-D than Avoid-F. Though there was no effect of DAT on ρ at test, 9-repeat carriers were on average fit with a higher value (Supplementary Table 2). It may be the case that this small parameter difference was enough to produce a behavioral effect in the absence of a significant difference in the parameter.

These results are also illuminating with respect to the influence of DAT on phasic and tonic DA (see section A Dopamine Genetic Composite Is Associated With Instructed Learning of the main text). Other striatal genes assayed in this paradigm asymmetrically affect approach and avoidance learning, as measured by genotypic differences in learning rates for gains and losses, respectively (Doll et al., 2011; Frank et al., 2007). Such differences are taken to reflect differences in the efficacy of phasic DA to affect learning, as learning rates govern the extent to which reward prediction errors conveyed by phasic DA update learned stimulus values. In the training phase, the finding that 9-carriers were best characterized as having an increased choice bias relative to 10/10 homozygotes is consistent with our hypothesis of an effect of DAT1 on tonic DA. In the test phase, however, the decreased learning rate for losses for 9-repeat carriers—in the absence of a significant difference in decision bias—is better explained by an effect on phasic DA. One potential way to reconcile these differences between training and test would be if

lower tonic DA in 9-repeat carriers produced less contrast in DA for the phasic dips thought to convey negative reward prediction errors (Niv, Duff, & Dayan, 2005).

2.3. DA Composite

There was a significant effect of DAC on ρ at training ($F(3,98) = 4.10, p = .009$), in line with the behavioral results. Post-hoc comparisons revealed that the DAC3 group had a significantly higher bias than all other groups (DAC3 vs. DAC0: $t(98) = 3.50, p_{corrected} = .004$; DAC3 vs. DAC1: $t(98) = 3.00, p_{corrected} = .017$; DAC3 vs. DAC2: $t(98) = 2.86, p_{corrected} = .02$). No other comparisons reached significance (all $p_{corrected} > .04$). There was also a significant gene-dose effect, with increasing DAC status associated with increasing bias ($r(100) = .26, p = .009$).

Again mirroring the Avoid-D/Avoid-F results, the effect of DAC on ρ during the test phase was not significant ($F(3,98) = 2.05, p = .11$). However, there was a significant gene-dose effect ($r(100) = .21, p = .03$).

2.4. tDCS

Though we found a significant effect of anodal stimulation during the training phase, there was no difference between the anodal and sham groups in the ρ parameter of the model ($t(65) = 0.54, p = .59$). Nor was there a difference between cathodal and sham ($t(68) = 0.57, p = .57$). Nor did we find differences in any other model parameters (all $ps > .36$). Because the anodal effect was only present early during training, we also refit the decision bias model on just the first two blocks of training phase data. We again found no difference in ρ between anodal and sham ($t(65) = -0.49, p = .63$). In keeping with the

behavioral results, we also found no difference in ρ during the test phase (Anodal vs. Sham: $t(65) = -0.56, p = .57$; Cathodal vs. Sham: $t(68) = 0.37, p = .71$).

In sum, we failed to find an effect of stimulation on model parameters. It may be that noise in the model parameter estimates prevented us from corroborating what was a very small behavioral effect at training for the comparison of anodal and sham. It may also be that anodal stimulation did not have a focal effect on any one parameter but rather induced weak, diffuse effects that together lead to a small behavioral difference in the absence of significant differences in model parameters (i.e., the numerically smaller learning rates and temperature parameters of the anodal group, combined with a numerically higher bias, could potentially have produced a small behavioral difference).

3. Supplementary Tables

3.1. Model Parameter Estimates

Supplementary Table 1. Parameter estimates for the decision bias model at training.

Group	N	α_g	α_l	β	ρ
Overall	103	0.34 (0.30)	0.21 (0.27)	0.29 (0.16)	0.16 (0.20)
COMT					
Val/Val	34	0.31 (0.30)	0.22 (0.30)	0.27 (0.17)	0.12 (0.18)
Val/Met	53	0.39 (0.31)	0.20 (0.26)	0.31 (0.17)	0.14 (0.20)
Met/Met	15	0.29 (0.28)	0.18 (0.24)	0.26 (0.15)	0.27 (0.19)
DAT					
10/10	67	0.34 (0.30)	0.20 (0.26)	0.30 (0.17)	0.12 (0.15)
9c	35	0.36 (0.31)	0.22 (0.29)	0.27 (0.16)	0.20 (0.25)
DA composite					
0	25	0.31 (0.29)	0.18 (0.26)	0.27 (0.18)	0.10 (0.13)
1	43	0.35 (0.30)	0.23 (0.29)	0.30 (0.16)	0.15 (0.20)
2	27	0.42 (0.32)	0.21 (0.28)	0.31 (0.17)	0.15 (0.20)
3	7	0.19 (0.12)	0.12 (0.17)	0.20 (0.11)	0.38 (0.22)
tDCS					
Anodal	33	0.29 (0.28)	0.20 (0.25)	0.26 (0.15)	0.17 (0.23)
Sham	34	0.35 (0.31)	0.21 (0.30)	0.29 (0.17)	0.14 (0.19)
Cathodal	36	0.39 (0.31)	0.21 (0.26)	0.30 (0.17)	0.16 (0.18)

Note. Parameter estimates are given as M (SD). Genotype counts only add up to 102 because genotyping failed for one subject.

Supplementary Table 2. Parameter estimates for the decision bias model at test.

Group	N	α_g	α_l	β	ρ
Overall	103	0.28 (0.35)	0.23 (0.33)	0.20 (0.08)	0.27 (0.31)
COMT					
Val/Val	34	0.24 (0.33)	0.18 (0.31)	0.19 (0.09)	0.19 (0.25)
Val/Met	53	0.32 (0.35)	0.28 (0.37)	0.20 (0.07)	0.29 (0.33)
Met/Met	15	0.26 (0.39)	0.19 (0.27)	0.19 (0.08)	0.37 (0.33)
DAT					
10/10	67	0.28 (0.33)	0.30 (0.36)	0.21 (0.08)	0.25 (0.30)
9c	35	0.29 (0.37)	0.11 (0.23)	0.18 (0.07)	0.32 (0.32)
DA composite					
0	25	0.26 (0.34)	0.20 (0.30)	0.19 (0.10)	0.19 (0.26)
1	43	0.26 (0.31)	0.33 (0.39)	0.20 (0.08)	0.27 (0.31)
2	27	0.36 (0.40)	0.15 (0.27)	0.22 (0.07)	0.28 (0.32)
3	7	0.20 (0.36)	0.10 (0.08)	0.13 (0.04)	0.51 (0.35)
tDCS					
Anodal	33	0.31 (0.36)	0.31 (0.39)	0.18 (0.07)	0.23 (0.29)
Sham	34	0.30 (0.33)	0.21 (0.30)	0.21 (0.09)	0.27 (0.29)
Cathodal	36	0.25 (0.35)	0.19 (0.31)	0.21 (0.08)	0.30 (0.34)

Note. Parameter estimates are given as M (SD). Genotype counts only add up to 102 because genotyping failed for one subject.

3.2. Demographics

Supplementary Table 3. Demographic breakdown of the 103 subjects included in the analyses after performance cutoffs.

Race/Ethnicity	N
Caucasian	50
Asian	24
African American	20
Other	9
Hispanic	
Y	14
N	87
Unknown	2

Supplementary Table 4. Demographic breakdown by tDCS condition.

	Anodal	Cathodal	Sham
Age			
Mean	21.82	21.92	21.76
SD	5.12	3.55	4.27
Gender			
M	10	17	11
F	23	19	23
Race			
Caucasian	15	23	12
African American	8	4	8
Asian	8	5	11
Other	2	4	3
Ethnicity			
Hispanic	4	5	5
Non-Hispanic	28	30	29
Genotype			
Val/Met	19	17	17
Met/Met	5	8	2
Val/Val	9	11	14
DAT 9c	12	14	9
DAT 10/10	21	22	24

Note. Genotype counts only add up to 102 because genotyping failed for one subject.

Supplementary Table 5. Demographic breakdown by genotype.

	COMT			DAT	
	Val/Met	Met/Met	Val/Vat	9c	10/10
Age					
Mean	22.70	21.80	20.62	21.09	22.28
SD	5.21	3.10	2.67	3.53	4.63
Gender					
M	24	6	7	14	23
F	29	9	27	21	44
Race					
Caucasian	26	13	10	21	28
African American	9	2	9	8	12
Asian	11	0	13	3	21
Other	7	0	2	3	6
Ethnicity					
Hispanic	7	1	6	7	7
Non-Hispanic	44	14	28	28	58
tDCS Condition					
Anodal	19	5	9	12	21
Cathodal	17	8	11	14	22
Sham	17	2	14	9	24

Note. Genotype counts only add up to 102 because genotyping failed for one subject.

Supplementary Table 6. Demographic breakdown by DA composite.

	DAC0	DAC1	DAC2	DAC3
Age				
Mean	21.20	22.21	22.15	21.14
SD	2.84	5.33	4.09	2.04
Gender				
M	4	17	15	1
F	21	26	12	6
Race				
Caucasian	6	19	18	6
African American	6	8	5	1
Asian	12	10	2	0
Other	1	6	2	0
Ethnicity				
Hispanic	5	3	5	1
Non-Hispanic	20	38	22	6
tDCS Condition				
Anodal	5	18	7	3
Cathodal	8	11	15	2
Sham	12	14	5	2

Note. Genotype counts only add up to 102 because genotyping failed for one subject.

3.3. Training Phase Results

Supplementary Table 7. ANOVA table for the mixed effects logistic regression model of the effect of COMT on instructed (CD vs. EF) training phase performance.

Predictor	χ^2	<i>df</i>	<i>p</i>
Intercept	6.24	1	.01
COMT	4.11	2	.13
Trial Type	66.24	1	< .0001
Block	14.78	3	.002
COMT x Trial Type	13.94	2	.0009
COMT x Block	2.74	6	.84
Trial Type x Block	0.68	3	.88
COMT x Trial Type x Block	7.83	6	.25

Note. Boldfaced text indicates $p < .05$.

Supplementary Table 8. ANOVA table for the mixed effects logistic regression model of the effect of COMT on uninstructed (AB, EF) training phase performance.

Predictor	χ^2	<i>df</i>	<i>p</i>
Intercept	194.47	1	< .0001
COMT	2.75	2	.25
Trial Type	76.99	1	< .0001
Block	34.39	3	< .0001
COMT x Trial Type	0.42	2	.81
COMT x Block	6.91	6	.33
Trial Type x Block	17.11	3	.0007

Note. Boldfaced text indicates $p < .05$.

Supplementary Table 9. Mixed effects logistic regression model of the effect of DAT on instructed (CD vs. EF) training phase performance.

Predictor	β	OR^a	z	p
Intercept	-0.32	0.73	-3.28	.001
9c vs. 10/10	-0.43	0.65	-2.17	.03
Trial Type	1.00	2.72	7.18	< .0001
Block 2 vs. 1	0.30	1.35	2.47	.01
Block 3 vs. (1,2)	0.25	1.28	2.33	.02
Block 4 vs. (1,2,3)	0.30	1.35	2.81	.005
9c vs. 10/10 x Trial Type	0.53	1.70	1.91	.06
9c vs. 10/10 x Block 2 vs. 1	-0.10	0.90	-0.41	.68
9c vs. 10/10 x Block 3 vs. (1,2)	0.12	1.13	0.56	.57
9c vs. 10/10 x Block 4 vs. (1,2,3)	0.02	1.02	0.08	.94
Trial Type x Block 2 vs. 1	-0.26	0.77	-1.72	.09
Trial Type x Block 3 vs. (1,2)	-0.02	0.98	-0.14	.89
Trial Type x Block 4 vs. (1,2,3)	-0.12	0.89	-0.77	.44
9c vs. 10/10 x Trial Type x Block 2 vs. 1	-0.24	0.79	-0.81	.42
9c vs. 10/10 x Trial Type x Block 3 vs. (1,2)	0.01	1.01	0.02	.98
9c vs. 10/10 x Trial Type x Block 4 vs. (1,2,3)	0.10	1.11	0.34	.73

Note. Boldfaced text indicates $p < .05$. ^a*OR*: Odds Ratio

Supplementary Table 10. Mixed effects logistic regression model of the effect of DAT on uninstructed (AB, EF) training phase performance.

Predictor	β	OR^a	z	p
Intercept	1.12	3.06	13.93	< .0001
9c vs. 10/10	-0.004	1.00	-0.03	.97
Trial Type	0.46	1.58	9.8	< .0001
Block 2 vs. 1	0.40	1.49	4.38	< .0001
Block 3 vs. (1,2)	0.23	1.26	2.77	.006
Block 4 vs. (1,2,3)	0.19	1.21	2.09	.04
9c vs. 10/10 x Trial Type	0.01	1.01	0.12	.91
9c vs. 10/10 x Block 2 vs. 1	-0.12	0.89	-0.73	.47
9c vs. 10/10 x Block 3 vs. (1,2)	0.11	1.12	0.65	.51
9c vs. 10/10 x Block 4 vs. (1,2,3)	0.03	1.03	0.16	.87
Trial Type x Block 2 vs. 1	0.33	1.39	4.10	< .0001
Trial Type x Block 3 vs. (1,2)	0.004	1.00	0.08	.94
Trial Type x Block 4 vs. (1,2,3)	0.02	1.02	0.27	.79

Note. Boldfaced text indicates $p < .05$. ^aOR: Odds Ratio

Supplementary Table 11. ANOVA table for the mixed effects logistic regression model of the effect of DAC on instructed (CD vs. EF) training phase performance.

Predictor	χ^2	df	p
Intercept	1.45	1	.23
DAC	11.03	3	.01
Trial Type	85.33	1	< .0001
Block	8.68	3	.03
DAC x Trial Type	29.61	3	< .0001
DAC x Block	9.68	9	.38
Trial Type x Block	2.10	3	.55
DAC x Trial Type x Block	9.12	9	.43

Note. Boldfaced text indicates $p < .05$.

Supplementary Table 12. ANOVA table for the mixed effects logistic regression model of the effect of DAC on uninstructed (AB, EF) training phase performance.

Predictor	χ^2	<i>df</i>	<i>p</i>
Intercept	174.86	1	< .0001
DAC	3.40	3	.33
Trial Type	67.10	1	< .0001
Block	26.87	3	< .0001
DAC x Trial Type	0.69	3	.88
DAC x Block	5.55	9	.78
Trial Type x Block	16.99	3	.0007

Note. Boldfaced text indicates $p < .05$.

Supplementary Table 13. Mixed effects logistic regression model of the effect of tDCS on uninstructed (AB, EF) training phase performance.

Predictor	β	<i>OR</i> ^a	<i>z</i>	<i>p</i>
Intercept	1.12	3.06	14.67	< .0001
Anodal vs. Sham	0.09	1.09	0.55	.58
Cathodal vs. Sham	-0.15	0.86	-0.88	.38
Trial Type	0.46	1.58	10.55	< .0001
Block 2 vs. 1	0.40	1.49	4.79	< .0001
Block 3 vs. (1,2)	0.22	1.25	2.74	.006
Block 4 vs. (1,2,3)	0.18	1.20	2.17	.03
Anodal vs. Sham x Trial Type	0.16	1.17	1.66	.10
Cathodal vs. Sham x Trial Type	-0.05	0.95	-0.48	.63
Anodal vs. Sham x Block 2 vs. 1	0.31	1.36	1.63	.10
Cathodal vs. Sham x Block 2 vs. 1	0.29	1.34	1.54	.12
Anodal vs. Sham x Block 3 vs. (1,2)	-0.06	0.94	-0.34	.74
Cathodal vs. Sham x Block 3 vs. (1,2)	-0.04	0.96	-0.19	.85
Anodal vs. Sham x Block 4 vs. (1,2,3)	0.07	1.07	0.38	.70
Cathodal vs. Sham x Block 4 vs. (1,2,3)	0.10	1.11	0.52	.60
Trial Type x Block 2 vs. 1	0.31	1.36	3.97	< .0001
Trial Type x Block 3 vs. (1,2)	0.02	1.02	0.26	.80
Trial Type x Block 4 vs. (1,2,3)	0.02	1.02	0.33	.74

Note. Boldfaced text indicates $p < .05$. ^a*OR*: Odds Ratio

Appendix B: Supplementary Material for Chapter 3

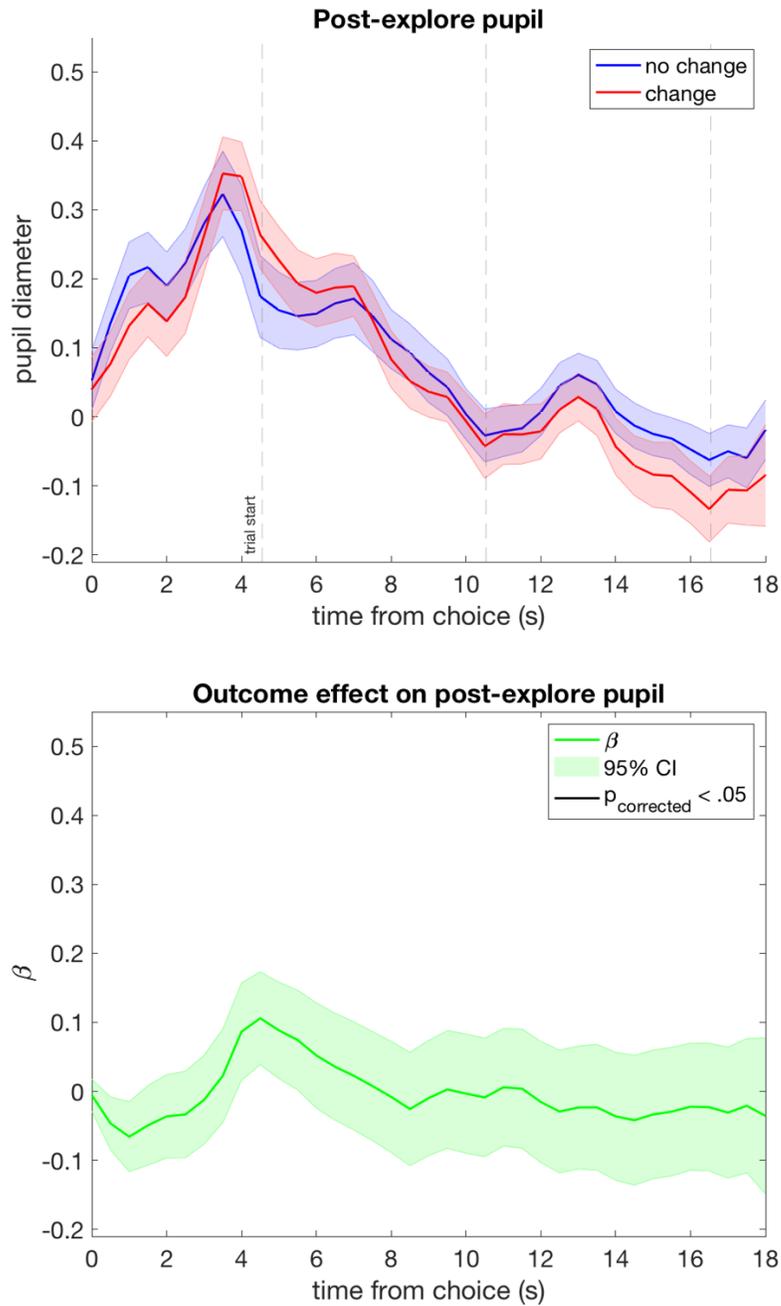


Figure S1. The effect of a change on post-explore pupil diameter: data (top) and a regression model controlling for gaze position (bottom). At no point was the effect of a change significant when controlling for multiple comparisons.

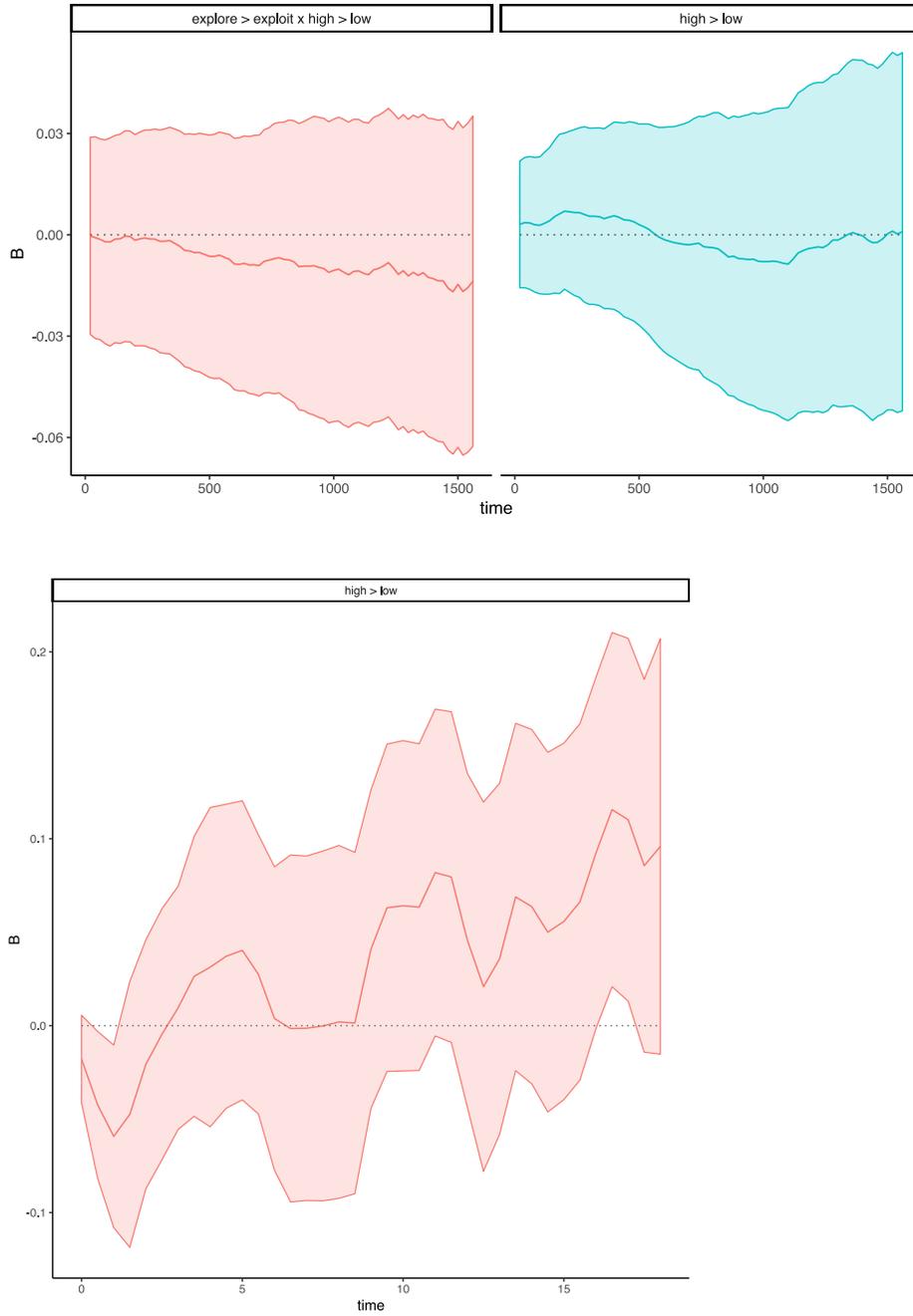


Figure S2. There was no significant effect of volatility condition on pupil diameter either at the trial level (top), or during the extended post-explore interval (bottom). In both cases, graphs show parameter estimates and 95% confidence intervals for the contrasts of interest.

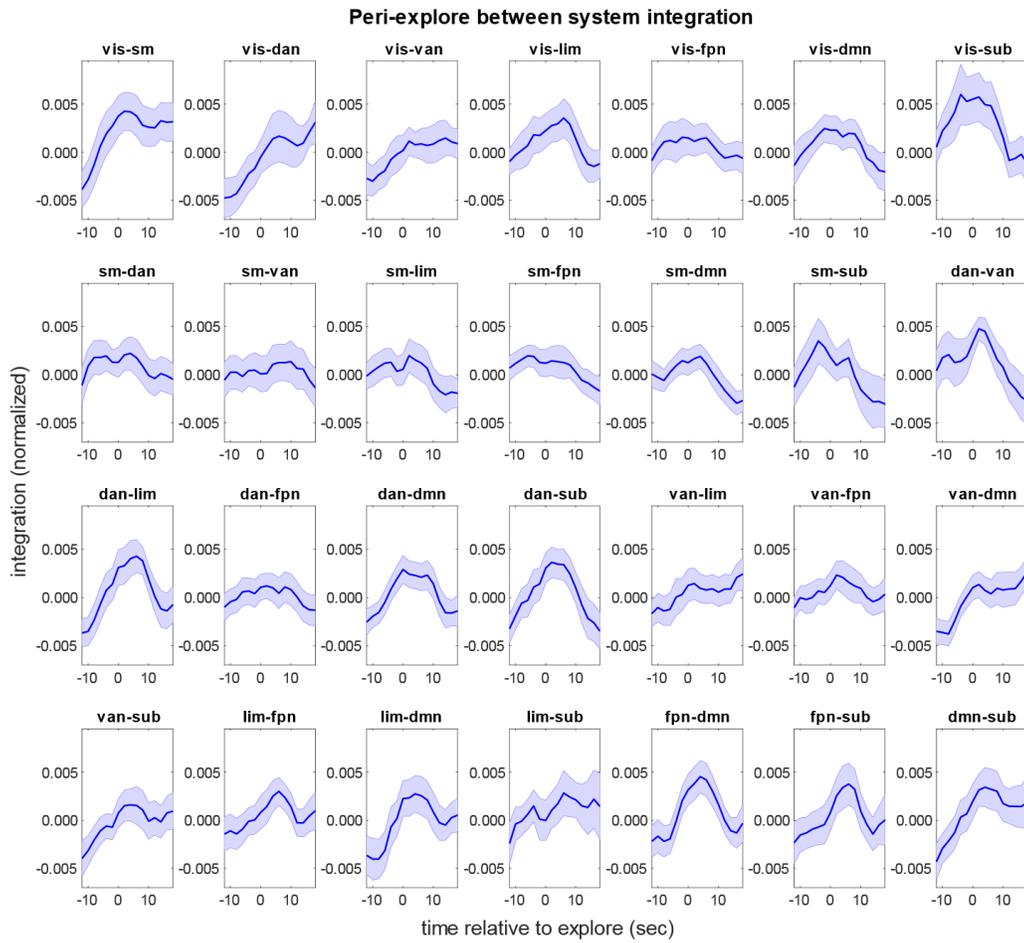


Figure S3. Peri-explore between-system integration for all pairs of cognitive systems.

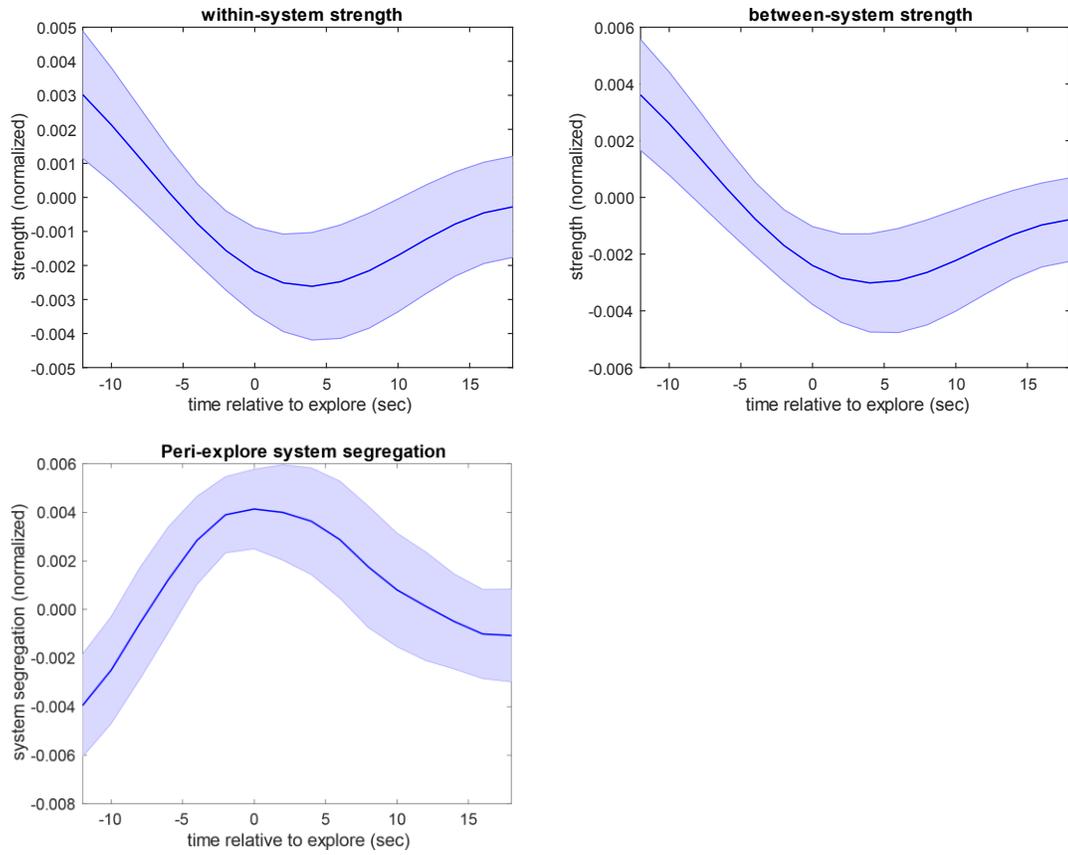


Figure S4. Peri-explore within and between system strength (top) and system segregation computed relative to the modules at each time point rather than to the Yeo cognitive systems (bottom).

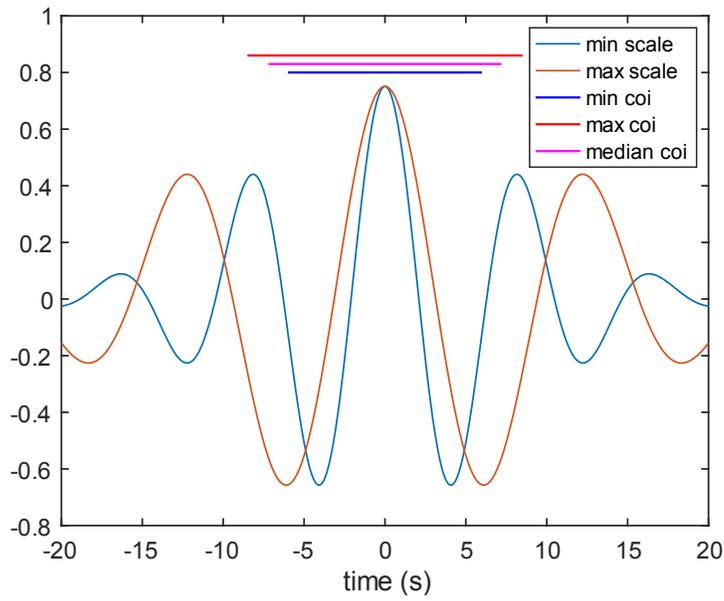


Figure S5. Visualization of the maximum and minimum wavelet scales used, in the real number domain. Thick bars above the wavelets indicate the width of the “cone of influence,” the central segment of the wavelet in which changes in the underlying signal have the greatest impact on wavelet power (Torrence & Compo, 1998).

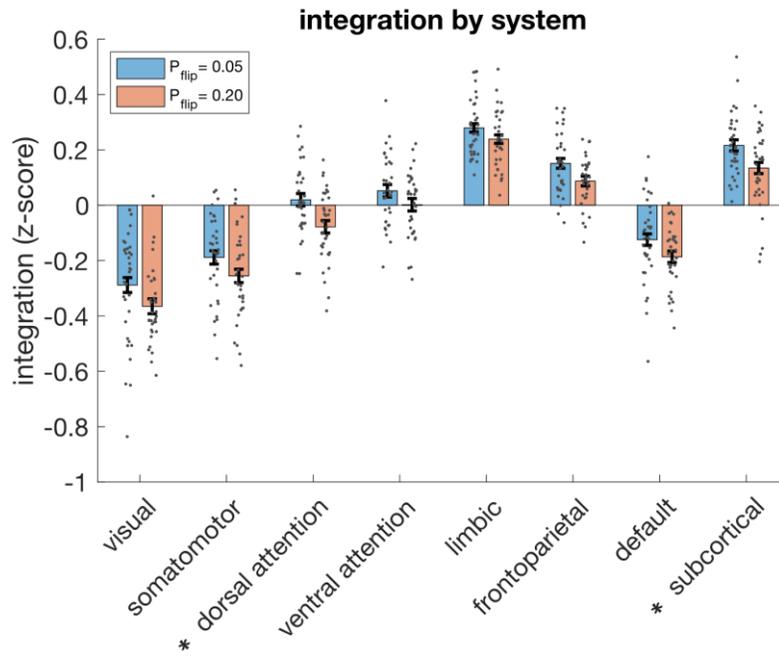


Figure S6. The effect of the volatility manipulation on integration for each cognitive system. * $p < .05$. Integration was z-scored within subject, across systems for visualization, but analyses were performed on untransformed values.

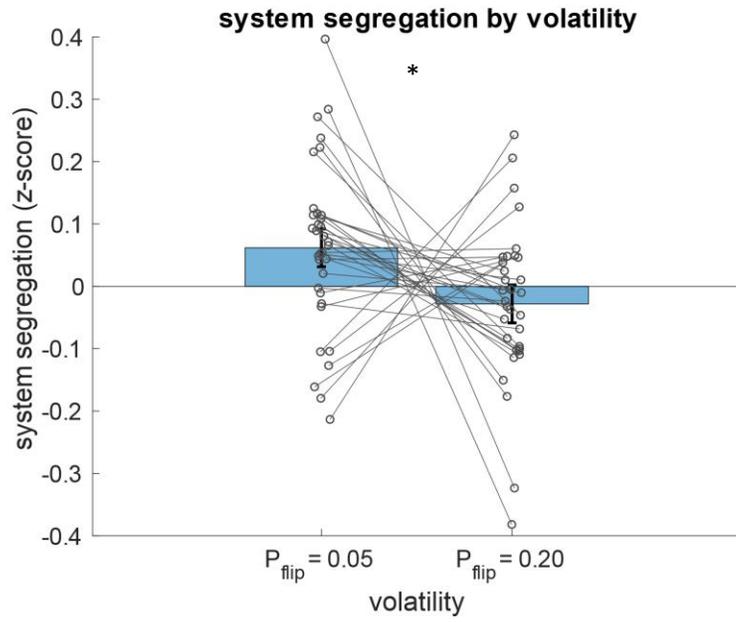


Figure S7. The effect of volatility on system segregation relative to the modules at each time point. * $p < .05$.

BIBLIOGRAPHY

- Aarts, E., Roelofs, A., Franke, B., Rijpkema, M., Fernández, G., Helmich, R. C., & Cools, R. (2010). Striatal dopamine mediates the interface between motivational and cognitive control in humans: Evidence from genetic imaging. *Neuropsychopharmacology*, *35*, 1943–1951.
- Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*.
<https://doi.org/10.1016/J.NEUROIMAGE.2019.05.011>
- Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*, 5369–5376.
- Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*, 1–1.
- Anzellotti, S., & Coutanche, M. N. (2018). Beyond Functional Connectivity: Investigating Networks of Multivariate Representations. *Trends in Cognitive Sciences*, *22*, 258–269.
- Arbabshirani, M. R., Damaraju, E., Phlypo, R., Plis, S., Allen, E., Ma, S., ... Calhoun, V. D. (2014). Impact of autocorrelation on functional connectivity. *NeuroImage*, *102*, 294–308.
- Arbuthnott, G. W., & Wickens, J. (2007). Space, time and dopamine. *Trends in*

Neurosciences, 30, 62–69.

Arnsten, A. F., Paspalas, C. D., Gamo, N. J., Yang, Y., & Wang, M. (2010). Dynamic Network Connectivity: A new form of neuroplasticity. *Trends in Cognitive Sciences*, 14, 365–375.

Arnsten, A. F. T. (2011). Catecholamine Influences on Dorsolateral Prefrontal Cortical Networks. *Biological Psychiatry*, 69, e89–e99.

Arsenault, J. T., Caspari, N., Vandenberghe, R., & Vanduffel, W. (2017). Attention shifts recruit the monkey default mode network. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 38, 1202–1217.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., ... Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, 526, 68–74.

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94, 206–234.

Bari, A., & Aston-Jones, G. (2013). Atomoxetine modulates spontaneous and sensory-evoked discharge of locus coeruleus noradrenergic neurons. *Neuropharmacology*, 64, 53–64.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and*

Language, 68, 255–278.

Bassett, D. S., & Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, 12, 512–523.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 7641–7646.

Bassett, D. S., Wymbs, N. F., Rombach, M. P., Porter, M. a, Mucha, P. J., & Grafton, S. T. (2013). Task-based core-periphery organization of human brain dynamics. *PLoS Computational Biology*, 9, e1003171.

Bassett, D. S., Yang, M., Wymbs, N. F., & Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*, 18, 744–751.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). *Parsimonious Mixed Models*. Retrieved from <http://arxiv.org/abs/1506.04967>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Bazzi, M., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2016). Community Detection in Temporal Multilayer Networks, with an Application to Correlation Networks. *Multiscale Modeling & Simulation*, 14, 1–41.

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, 20, 87–95.

Beeler, J. A., Daw, N., Frazier, C. R. M., & Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in Behavioral Neuroscience*, 4,

170.

- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37, 90–101.
- Bendesky, A., Tsunozaki, M., Rockman, M. V., Kruglyak, L., & Bargmann, C. I. (2011). Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature*, 472, 313–318.
- Bernatskiy, A., & Bongard, J. C. (2015). Exploiting the relationship between structural modularity and sparsity for faster network evolution. *GECCO 2015 - Companion Publication of the 2015 Genetic and Evolutionary Computation Conference*, 1173–1176. Association for Computing Machinery, Inc.
- Bernstein, E., Shore, J., & Lazer, D. (2018). How intermittent breaks in interaction improve collective intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, 201802407.
- Berridge, C. W., & Waterhouse, B. D. The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. , 42 *Brain Research Reviews* § (2003).
- Berridge, C. W., & Waterhouse, B. D. (2003b). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42, 33–84.
- Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *European Journal of Neuroscience*, 35, 1124–1143.

- Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., & D'Esposito, M. (2018, October). A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour*, Vol. 2, pp. 765–777. Nature Publishing Group.
- Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science*, 33, 206–242.
- Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biology*, 9, e1001089.
- Bilder, R. M., Volavka, J., Lachman, H. M., & Grace, A. A. (2004, November 11). The catechol-O-methyltransferase polymorphism: Relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. *Neuropsychopharmacology*, Vol. 29, pp. 1943–1961. Nature Publishing Group.
- Birn, R. M., Converse, A. K., Rajala, A. Z., Alexander, A. L., Block, W. F., McMillan, A. B., ... Populin, L. C. (2019). Changes in Endogenous Dopamine Induced by Methylphenidate Predict Functional Connectivity in Nonhuman Primates. *The Journal of Neuroscience*, 39, 1436–1444.
- Blackwell, K. A., Chatham, C. H., Wiseheart, M., & Munakata, Y. (2014). A developmental window into trade-offs in executive function: The case of task switching versus response inhibition in 6-year-olds. *Neuropsychologia*, 62, 356–364.
- Blanco, N. J., Love, B. C., Cooper, J. A., McGeary, J. E., Knopik, V. S., & Maddox, W. T. (2015). A frontal dopamine system for reflective exploratory behavior. *Neurobiology of Learning and Memory*, 123, 84–91.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and*

Experiment, 2008, P10008.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, 8.

<https://doi.org/10.1038/ncomms15958>

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001).

Conflict Monitoring and Cognitive Control. *Psychological Review*, 108, 624–652.

Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, 28, 574–582.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., ...

Bassett, D. S. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 11678–11683.

Brisch, R., Saniotis, A., Wolf, R., Biela, H., Bernstein, H.-G., Steiner, J., ... Gos, T.

(2014). The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: Old fashioned, but still in vogue. *Frontiers in Psychiatry*, 5, 47.

Cagniard, B., Beeler, J. A., Britt, J. P., McGehee, D. S., Marinelli, M., & Zhuang, X.

(2006). Dopamine scales performance in the absence of new learning. *Neuron*, 51, 541–547.

Cattaneo, Z., Pisoni, A., & Papagno, C. (2011). Transcranial direct current stimulation over Broca's region improves phonemic and semantic fluency in healthy

individuals. *Neuroscience*, 183, 64–70.

Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E., & Wig, G. S. (2014). Decreased

- segregation of brain systems across the healthy adult lifespan. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E4997–E5006.
- Chrysikou, E. G., Hamilton, R. H., Coslett, H. B., Datta, A., Bikson, M., & Thompson-Schill, S. L. (2013). Noninvasive transcranial direct current stimulation over the left prefrontal cortex facilitates cognitive flexibility in tool use. *Cognitive Neuroscience*, *4*, 81–89.
- Chrysikou, E. G., Weber, M. J., & Thompson-Schill, S. L. (2014). A matched filter hypothesis for cognitive control. *Neuropsychologia*, *62*, 341–355.
- Clewett, D., & Murty, V. (2019). Echoes of emotions past: How neuromodulators determine what we recollect. *Eneuro*, ENEURO.0108-18.2019.
- Clune, J., Mouret, J.-B., & Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20122863.
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*, 933–942.
- Cohen, J. R., & D’Esposito, M. (2016). The Segregation and Integration of Distinct Brain Networks and Their Relationship to Cognition. *Journal of Neuroscience*, *36*.
- Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *The Journal of Neuroscience*, *37*, 4332–4342.
- Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*, 1024–1035.

- Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*, 337–366.
- Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biology*, *10*, e1001293.
- Cools, R., & D'Esposito, M. (2009). Dopaminergic modulation of flexible cognitive control in humans. In L. L. Iversen, S. D. Iversen, S. B. Dunnett, & A. Bjorklund (Eds.), *Dopamine handbook* (pp. 249–260). New York, NY: Oxford University Press.
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*, 306–324.
- Costa, A., Riedel, M., Müller, U., Möller, H.-J., & Ettinger, U. (2011). Relationship between SLC6A3 genotype and striatal dopamine transporter availability: A meta-analysis of human single photon emission computed tomography studies. *Synapse*, *65*, 998–1005.
- Coutanche, M. N., & Thompson-Schill, S. L. (2013). Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Frontiers in Human Neuroscience*, *7*. <https://doi.org/10.3389/fnhum.2013.00015>
- Cragg, S. J., & Rice, M. E. (2004). DANCING past the DAT at a DA synapse. *Trends in Neurosciences*, *27*, 270–277.
- Crittenden, B. M., Mitchell, D. J., Duncan, J., Cohen, J., Andrews, J., Raichle, M., ... Johnson, K. (2015). Recruitment of the default mode network during a demanding act of executive control. *ELife*, *4*, e06481.

- Cruzat, J., Deco, G., Tauste-Campo, A., Principe, A., Costa, A., Kringelbach, M. L., & Rocamora, R. (2018). The dynamics of human cognition: Increasing global integration coupled with decreasing segregation found using iEEG. *NeuroImage*, *172*, 492–505.
- da Silva Alves, F., Figeo, M., van Amelsvoort, T., Veltman, D., & de Haan, L. (2008). The revised dopamine hypothesis of schizophrenia: Evidence from pharmacological MRI studies with atypical antipsychotic medication. *Psychopharmacology Bulletin*, *41*, 121–132.
- Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*, e1003441.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.
- Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.
- de Cheveigné, A., & Nelken, I. (2019). Filters: When, Why, and How (Not) to Use Them. *Neuron*, *102*, 280–293.
- Dean, R. T., & Dunsmuir, W. T. M. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, *48*, 783–802.
- Deco, G., Tononi, G., Boly, M., & Kringelbach, M. L. (2015). Rethinking segregation

- and integration: contributions of whole-brain modelling. *Nature Reviews Neuroscience*, *16*, 430–439.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611–1618.
- den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., ... Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, *80*, 1090–1100.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.
- Dixon, M. L., Andrews-Hanna, J. R., Spreng, R. N., Irving, Z. C., Mills, C., Girn, M., & Christoff, K. (2017). Interactions between the default network and dorsal attention network vary across default subsystems, time, and cognitive states. *NeuroImage*, *147*, 632–649.
- Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E1598–E1607.
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*, 312–325.
- Doll, B. B., Bath, K. G., Daw, N. D., & Frank, M. J. (2016). Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. *Journal of Neuroscience*, *36*, 1211–1222.
- Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *Journal of*

- Neuroscience*, 31, 6188–6198.
- Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research*, 1299, 74–94.
- Doll, B. B., Waltz, J. A., Cockburn, J., Brown, J. K., Frank, M. J., & Gold, J. M. (2014). Reduced susceptibility to confirmation bias in schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience*, 14, 715–728.
- Doucette-Stamm, L. A., Blakely, D. J., Tian, J., Mockus, S., & Mao, J. (1995). Population genetic study of the human dopamine transporter gene (DAT1). *Genetic Epidemiology*, 12, 303–308.
- Dreher, J.-C., Kohn, P., Kolachana, B., Weinberger, D. R., & Berman, K. F. (2009). Variation in dopamine genes influences responsivity of the human reward system. *Proceedings of the National Academy of Sciences*, 106, 617–622.
- Dreyer, J. K., Herrik, K. F., Berg, R. W., & Hounsgaard, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. *Journal of Neuroscience*, 30, 14273–14283.
- Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-O-methyltransferase genotypes and schizophrenia. *Biological Psychiatry*, 64, 739–749.
- Ebitz, R. B., Albarran, E., & Moore, T. (2018). Exploration Disrupts Choice-Predictive Signals and Alters Dynamics in Prefrontal Cortex. *Neuron*, 97, 1–12.
- Ebitz, R. B., & Moore, T. (2016). Altered balance between top-down and bottom-up saccade control across exploration and exploitation. *Gordon Research Seminar*.

- Newry, M.E.
- Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*, 85, 628–640.
- Efimova, E. V., Gainetdinov, R. R., Budygin, E. A., & Sotnikova, T. D. (2016). Dopamine transporter mutant animals: A translational perspective. *Journal of Neurogenetics*, 30, 5–15.
- Ekman, M., Derrfuss, J., Tittgemeyer, M., & Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences*, 109, 16714–16719.
- Elber-Dorozko, L., & Loewenstein, Y. (2018). Striatal action-value neurons reconsidered. *ELife*, 7. <https://doi.org/10.7554/eLife.34248>
- Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, 16, 1146–1153.
- Eldar, E., Niv, Y., & Cohen, J. D. (2016). Do You See the Forest or the Tree? Neural Gain and Breadth Versus Focus in Perceptual Processing. *Psychological Science*, 27, 1632–1643.
- Faraone, S. V., Spencer, T. J., Madras, B. K., Zhang-James, Y., & Biederman, J. (2014). Functional effects of dopamine transporter gene genotypes on in vivo dopamine transporter functioning: A meta-analysis. *Molecular Psychiatry*, 19, 880–889.
- Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. *Frontiers in Neuroscience*, 6, 148.
- Fay, M. P., & Shaw, P. A. (2010). Exact and asymptotic weighted logrank tests for

- interval censored data: The interval R package. *Journal of Statistical Software*, 36.
<https://doi.org/10.18637/jss.v036.i02>
- Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, 18, 120–126.
- Filmer, H. L., Mattingley, J. B., & Dux, P. E. (2013). Improved multitasking following prefrontal tDCS. *Cortex*, 49, 2845–2852.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299, 1898–1902.
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62, 774–781.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., ... Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X)
- Floresco, S. B., West, A. R., Ash, B., Moore, H., & Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature Neuroscience*, 6, 968–973.
- Ford, C. P., Gantz, S. C., Phillips, P. E. M., & Williams, J. T. (2010). Control of extracellular dopamine at dendrite and axon terminals. *Journal of Neuroscience*, 30, 6975–6983.
- Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., & Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *Journal of Neuroscience*, 33, 3602–3611.
- Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd ed.).

Thousand Oaks, CA: Sage.

- Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience, 12*, 1062–1068.
- Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, 104*, 16311–16316.
- Franke, B., Vasquez, A. A., Johansson, S., Hoogman, M., Romanos, J., Boreatti-Hümmer, A., ... Reif, A. (2010). Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD. *Neuropsychopharmacology, 35*, 656–664.
- Fregni, F., Boggio, P. S., Nitsche, M., Berman, F., Antal, A., Feredoes, E., ... Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research, 166*, 23–30.
- Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology, 34*, 475–484.
- Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series. *Magnetic Resonance in Medicine, 35*, 346–355.
- Gallagher, M., & Holland, P. C. (1994). The amygdala complex: multiple roles in

- associative learning and attention. *Proceedings of the National Academy of Sciences*, *91*, 11771–11776.
- Gallos, L. K., Makse, H. A., & Sigman, M. (2012). A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 2825–2830.
- Garcia-Garcia, M., Barceló, F., Clemente, I. C., & Escera, C. (2010). The role of the dopamine transporter DAT1 genotype on the neural correlates of cognitive flexibility. *European Journal of Neuroscience*, *31*, 754–760.
- Gerraty, R. T., Davidow, J. Y., Foerde, K., Galvan, A., Bassett, D. S., & Shohamy, D. (2018). Dynamic Flexibility in Striatal-Cortical Circuits Supports Reinforcement Learning. *The Journal of Neuroscience*, *38*, 2442–2453.
- Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, *71*, 1–6.
- Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.
- Giessing, C., Thiel, C. M., Alexander-Bloch, A. F., Patel, A. X., & Bullmore, E. T. (2013). Human brain functional network changes associated with enhanced and impaired attentional task performance. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*, 5903–5914.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*, 252–269.

- Good, B. H., De Montjoye, Y. A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 81. <https://doi.org/10.1103/PhysRevE.81.046106>
- Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, 24, 87–92.
- Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience*, 8, 805–812.
- Grace, A. A., & Gomes, F. V. (2018). The circuitry of dopamine system regulation and its disruption in schizophrenia: Insights into treatment and prevention. *Schizophrenia Bulletin*, sbx199. <https://doi.org/10.1093/schbul/sbx199>
- Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2, 111–120.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48, 63–72.
- Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11, 561–566.
- Gruner, P., & Pittenger, C. (2017, March 14). Cognitive inflexibility in Obsessive-Compulsive Disorder. *Neuroscience*, Vol. 345, pp. 243–255. Elsevier Ltd.
- Guedj, C., Monfardini, E., Reynaud, A. J., Farnè, A., Meunier, M., & Hadj-Bouziane, F. (2016). Boosting Norepinephrine Transmission Triggers Flexible Reconfiguration of Brain Networks at Rest. *Cerebral Cortex (New York, N.Y. : 1991)*.

<https://doi.org/10.1093/cercor/bhw262>

- Halassa, M. M., & Kastner, S. (2017). Thalamic functions in distributed cognitive control. *Nature Neuroscience*, *20*, 1669–1679.
- Hasselmo, M. E., Linster, C., Patil, M., Ma, D., & Cekic, M. (1997). Noradrenergic suppression of synaptic transmission may influence cortical signal-to-noise ratio. *Journal of Neurophysiology*, *77*, 3326–3339.
- Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *Journal of Neuroscience*, *31*, 4178–4187.
- Hayes, T. R., & Petrov, A. A. (2016). Pupil Diameter Tracks the Exploration–Exploitation Trade-off during Analogical Reasoning and Explains Individual Differences in Fluid Intelligence. *Journal of Cognitive Neuroscience*, *28*, 308–318.
- Heilbronner, S. R., & Platt, M. L. (2013). Causal evidence of performance monitoring by neurons in posterior cingulate cortex during learning. *Neuron*, *80*, 1384–1391.
- Herd, S. A., O'Reilly, R. C., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia*, *62*, 375–389.
- Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, *29*, 188–209.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192.
- Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a

- quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*. <https://doi.org/10.3758/BF03204445>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.
- Honari, H., Choe, A. S., Pekar, J. J., & Lindquist, M. A. (2019). Investigating the impact of autocorrelation on time-varying connectivity. *NeuroImage*. <https://doi.org/10.1016/J.NEUROIMAGE.2019.04.042>
- Hurley, L., Devilbiss, D., & Waterhouse, B. (2004). A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. *Current Opinion in Neurobiology*, *14*, 488–495.
- Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: A meta-analytical review. *Experimental Brain Research*, *216*, 1–10.
- Jenkinson, M. (2003). Fast, automated, N-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, *49*, 193–197.
- Jenkinson, M. (2004). Improving the registration of B0-distorted EPI images using calculated cost function weights. *Tenth International Conference on Functional Mapping of the Human Brain*.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*, 782–790.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*, 1587–1596.

- Jepma, M., Te Beek, E. T., Wagenmakers, E.-J., van Gerven, J. M. A., & Nieuwenhuis, S. (2010). The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Frontiers in Human Neuroscience*, *4*, 170.
- Jeub, L. G. S., Bazzi, M., Jutla, I. S., & Mucha, P. J. (2011). *A generalized Louvain method for community detection implemented in MATLAB*.
- Jiang, J., Beck, J., Heller, K., & Egner, T. (2015). *An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands*. <https://doi.org/10.1038/ncomms9165>
- Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience and Biobehavioral Reviews*, *46*, 30–43.
- Jo, H. J., Saad, Z. S., Simmons, W. K., Milbury, L. A., & Cox, R. W. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, *52*, 571–582.
- Johansen, S. (2012). *The Analysis of Nonstationary Time Series Using Regression, Correlation and Cointegration*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2179946
- Joshi, S., Li, Y., Kalwani, R. M. M., & Gold, J. I. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*, 221–234.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science (New York, N.Y.)*, *154*, 1583–1585.

- Kane, G. A., Vazey, E. M., Wilson, R. C., Shenhav, A., Daw, N. D., Aston-Jones, G., & Cohen, J. D. (2017). *Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task*. <https://doi.org/10.3758/s13415-017-0531-y>
- Karahanoğlu, F. I., Caballero-Gaudes, C., Lazeyras, F., & Van De Ville, D. (2013). Total activation: fMRI deconvolution through spatio-temporal regularization. *NeuroImage*, *73*, 121–134.
- Karuza, E. A., Balewski, Z. Z., Hamilton, R. H., Medaglia, J. D., Tardiff, N., & Thompson-Schill, S. L. (2016). Mapping the parameter space of tDCS and cognitive control via manipulation of current polarity and intensity. *Frontiers in Human Neuroscience*, *10*. <https://doi.org/10.3389/fnhum.2016.00665>
- Kaya, C., Cheng, M. H., Block, E. R., Bartol, T. M., Sejnowski, T. J., Sorkin, A., ... Bahar, I. (2018). Heterogeneities in axonal structure and transporter distribution lower dopamine reuptake efficiency. *ENeuro*, *5*, ENEURO.0298-17.2017.
- Khambhati, A. N., Mattar, M. G., Wymbs, N. F., Grafton, S. T., & Bassett, D. S. (2018). Beyond modularity: Fine-scale mechanisms and rules for brain network reconfiguration. *NeuroImage*, *166*, 385–399.
- King, D., Zigmond, M. J., & Finlay, J. M. (1997). Effects of dopamine depletion in the medial prefrontal cortex on the stress-induced increase in extracellular dopamine in the nucleus accumbens core and shell. *Neuroscience*, *77*, 141–153.
- Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, *3*, 398.
- Kohno, M., Nurmi, E. L., Laughlin, C. P., Morales, A. M., Gail, E. H., Helleman, G. S., & London, E. D. (2016). Functional genetic variation in dopamine signaling

moderates prefrontal cortical activity during risky decision making.

Neuropsychopharmacology, 41, 695–703.

Kolachana, B. S., Saunders, R. C., & Weinberger, D. R. (1995). Augmentation of prefrontal cortical monoaminergic activity inhibits dopamine release in the caudate nucleus: An in vivo neurochemical assessment in the rhesus monkey. *Neuroscience*, 69, 859–868.

Krugel, L. K., Biele, G., Mohr, P. N. C., Li, S.-C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences*, 106, 17951–17956.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674–689.

Lavín, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, 7. <https://doi.org/10.3389/fnbeh.2013.00218>

Lazer, D., & Friedman, A. (2007). The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52, 667–694.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, 69, 1–33.

Leonardi, N., & Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, 104, 430–436.

Levin, M. E., MacLane, C., Daflos, S., Seeley, J. R., Hayes, S. C., Biglan, A., & Pistorello, J. (2013). Examining psychological inflexibility as a transdiagnostic

- process across psychological disorders. *Journal of Contextual Behavioral Science*, *3*, 155–163.
- Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences*, *108*, 55–60.
- Liu, X., Zhang, N., Chang, C., & Duyn, J. H. (2018, October 15). Co-activation patterns in resting-state fMRI signals. *NeuroImage*, Vol. 180, pp. 485–494. Academic Press Inc.
- Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition*, *124*, 36–49.
- MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*, 161–175.
- Mason, W. A., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, *137*, 422–433.
- Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*, 764–769.
- Mattar, M. G., Cole, M. W., Thompson-Schill, S. L., & Bassett, D. S. (2015). A Functional Cartography of Cognitive Systems. *PLOS Computational Biology*, *11*, e1004533.
- McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., ...

- McCormick, D. A. (2015). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron*, *87*, 1143–1161.
- Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of Cognitive Neuroscience*, *27*, 1471–1491.
- Medaglia, J. D., Satterthwaite, T. D., Kelkar, A., Ciric, R., Moore, T. M., Ruparel, K., ... Bassett, D. S. (2018). Brain state expression and transitions are related to complex executive cognition in normative neurodevelopment. *NeuroImage*, *166*, 293–306.
- Medic, N., Ziauddeen, H., Vestergaard, M. D., Henning, E., Schultz, W., Farooqi, I. S., & Fletcher, P. C. (2014). Dopamine modulates the neural representation of subjective value of food in hungry subjects. *Journal of Neuroscience*, *34*, 16856–16864.
- Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., ... Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215.
- Meyer-Lindenberg, A., Kohn, P. D., Kolachana, B., Kippenhan, S., McInerney-Leo, A., Nussbaum, R., ... Berman, K. F. (2005). Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nature Neuroscience*, *8*, 594–596.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Minarik, T., Berger, B., Althaus, L., Bader, V., Biebl, B., Brotzeller, F., ... Sauseng, P. (2016). The importance of sample size for reproducibility of tDCS effects. *Frontiers in Human Neuroscience*, *10*, 453.
- Mohanty, A., Gitelman, D. R., Small, D. M., & Mesulam, M. M. (2008). The spatial

- attention network interacts with limbic and monoaminergic systems to modulate motivation-induced attention shifts. *Cerebral Cortex*, *18*, 2604–2613.
- Monti, M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Frontiers in Human Neuroscience*, *5*, 28.
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science*, *328*, 876–878.
- Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing Cognitive Control. *Current Directions in Psychological Science*, *21*, 71–77.
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*, 1040–1046.
- Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*, 12366–12378.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.
- Nieratschker, V., Kiefer, C., Giel, K., Krüger, R., & Plewnia, C. (2015). The COMT Val/Met polymorphism modulates effects of tDCS on response inhibition. *Brain Stimulation*, *8*, 283–288.
- Nikolova, Y. S., Ferrell, R. E., Manuck, S. B., & Hariri, A. R. (2011). Multilocus genetic profile for dopamine signaling predicts ventral striatum reactivity.

- Neuropsychopharmacology*, 36, 1940–1947.
- Nitsche, M. A., Nitsche, M. S., Klein, C. C., Tergau, F., Rothwell, J. C., & Paulus, W. (2003). Level of action of cathodal DC polarisation induced inhibition of the human motor cortex. *Clinical Neurophysiology*, 114, 600–604.
- Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*, 527, 633–639.
- Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, 57, 1899–1901.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53, 139–154.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191, 507–520.
- Niv, Y., Duff, M. O., & Dayan, P. D. (2005). Dopamine, uncertainty and TD learning. *Behavioral and Brain Functions*, 1, 6.
- Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, 12, 265–272.
- Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt? *Neuropsychologia*, 51, 2770–2780.
- Nozari, N., Woodard, K., & Thompson-Schill, S. L. (2014). Consequences of cathodal stimulation for behavior: When does it help and when does it hurt performance? *PLoS ONE*, 9, e84338.
- O’Doherty, J. P. (2004). Reward representations and reward-related learning in the

- human brain: Insights from neuroimaging. *Current Opinion in Neurobiology*, *14*, 769–776.
- O'Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. [Http://Dx.Doi.Org/10.1162/089976606775093909](http://dx.doi.org/10.1162/089976606775093909).
- Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013a). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, *24*, 751–761.
- Otto, A. R., Knox, W. B., Markman, A. B., & Love, B. C. (2014). Physiological and behavioral signatures of reflective exploratory choice. *Cognitive, Affective & Behavioral Neuroscience*, *14*, 1167–1183.
- Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013b). Working-memory capacity protects model-based learning from stress. *110*.
<https://doi.org/10.1073/pnas.1312011110>
- Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*, *27*, 319–333.
- Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLoS Computational Biology*, *7*, e1001048.
- Pearson, J. M., Hayden, B. Y., Raghavachari, S., & Platt, M. L. (2009). Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Current Biology : CB*, *19*, 1532–1537.

- Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in Cognitive Sciences, 15*, 143–151.
- Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ, 7*, e6876.
- Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics, 33*, 311–340.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). *{nlme}: Linear and Nonlinear Mixed Effects Models*. Retrieved from <https://cran.r-project.org/package=nlme>
- Plewnia, C., Zwissler, B., Längst, I., Maurer, B., Giel, K., & Krüger, R. (2013). Effects of transcranial direct current stimulation (tDCS) on executive functions: Influence of COMT Val/Met polymorphism. *Cortex, 49*, 1801–1807.
- Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*. <https://doi.org/10.3389/fnins.2011.00115>
- Pyper, B. J., & Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences, 55*, 2127–2140.
- Qiyuan, J., Richer, F., Wagoner, B. L., & Beatty, J. (1985). The Pupil and Stimulus Probability. *Psychophysiology, 22*, 530–534.

- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, 38, 433–447.
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for Neuroscience Abstracts*, 19, 974.
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7, 13289.
- Remijnse, P. L., van den Heuvel, O. A., Nielen, M. M. A., Vriend, C., Hendriks, G. J., Hoogendijk, W. J. G., ... Veltman, D. J. (2013). Cognitive Inflexibility in Obsessive-Compulsive Disorder and Major Depression Is Associated with Distinct Neural Correlates. *PLoS ONE*, 8. <https://doi.org/10.1371/journal.pone.0059600>
- Rice, M. E., & Cragg, S. J. (2008). Dopamine spillover after quantal release: Rethinking dopamine transmission in the nigrostriatal pathway. *Brain Research Reviews*, 58, 303–313.
- Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - Revisited. *NeuroImage*, 84, 971–985.
- Robbins, T. W., & Arnsten, A. F. T. (2009). The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annual Review of Neuroscience*,

32, 267–287.

- Roffman, J. L., Tanner, A. S., Eryilmaz, H., Rodriguez-Thompson, A., Silverstein, N. J., Ho, N. F., ... Catana, C. (2016). Dopamine D1 signaling organizes network dynamics underlying working memory. *Science Advances*, 2.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52, 1059–1069.
- Sadleir, R. J., Vannorsdall, T. D., Schretlen, D. J., & Gordon, B. (2010). Transcranial direct current stimulation (tDCS) in a realistic head model. *NeuroImage*, 51, 1310–1318.
- Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology*, 191, 461–482.
- Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron*, 76, 130–141.
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., ... Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, 28, 3095–3114.
- Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20, 416–420.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, 74, 1–58.

- Shafiei, G., Zeighami, Y., Clark, C. A., Coull, J. T., Nagano-Saito, A., Leyton, M., ...
Mišić, B. (2019). Dopamine Signaling Modulates the Stability and Integration of
Intrinsic Brain Networks. *Cerebral Cortex*, *29*, 397–409.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control:
An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, *79*, 217–240.
- Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering.
Psychobiology, *28*, 207–218.
- Shine, J. M., Aburn, M. J., Breakspear, M., & Poldrack, R. A. (2018a). The modulation
of neural gain facilitates a transition between functional segregation and integration
in the brain. *ELife*, *7*, e31130.
- Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., ...
Poldrack, R. A. (2016). The Dynamics of Functional Brain Networks: Integrated
Network States during Cognitive Task Performance. *Neuron*, *0*, 233–246.
- Shine, J. M., van den Brink, R. L., Hernaus, D., Nieuwenhuis, S., & Poldrack, R. A.
(2018b). Catecholaminergic manipulation alters dynamic network topology across
cognitive states. *Network Neuroscience*, *2*, 381–396.
- Shiner, T., Seymour, B., Wunderlich, K., Hill, C., Bhatia, K. P., Dayan, P., & Dolan, R. J.
(2012). Dopamine and performance in a reinforcement learning task: Evidence from
Parkinson's disease. *Brain*, *135*, 1871–1883.
- Shore, J., Bernstein, E., & Lazer, D. (2015). Facts and figuring: An experimental
investigation of network structure and performance in information and solution
spaces. *Organization Science*, *26*, 1432–1446.
- Shumay, E., Chen, J., Fowler, J. S., & Volkow, N. D. (2011). Genotype and ancestry

- modulate brain's DAT availability in healthy humans. *PLoS ONE*, 6, e22754.
- Slifstein, M., van de Giessen, E., Van Snellenberg, J., Thompson, J. L., Narendran, R., Gil, R., ... Abi-Dargham, A. (2015). Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia: A positron emission tomographic functional magnetic resonance imaging study. *JAMA Psychiatry*, 72, 316.
- Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Research*, 1428, 60–70.
- Smittenaar, P., Chase, H. W., Aarts, E., Nusslein, B., Bloem, B. R., & Cools, R. (2012). Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection - learning or performance? *European Journal of Neuroscience*, 35, 1144–1151.
- Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, 80, 914–919.
- Staudinger, M. R., & Büchel, C. (2013). How initial confirmatory experience potentiates the detrimental influence of bad advice. *NeuroImage*, 76, 125–133.
- Sulzer, D., Cragg, S. J., & Rice, M. E. (2016). Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia*, 6, 123–148.
- Sun, F. T., Miller, L. M., & D'Esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *NeuroImage*, 21, 647–658.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.).

Cambridge, MA: MIT Press.

- Tan, H.-Y., Chen, Q., Sust, S., Buckholtz, J. W., Meyers, J. D., Egan, M. F., ... Callicott, J. H. (2007). Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes on working memory brain function. *Proceedings of the National Academy of Sciences, 104*, 12536–12541.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society, 79*, 61–78.
- Totah, N. K. B., Logothetis, N. K., & Eschenko, O. Noradrenergic ensemble-based modulation of cognition over multiple timescales. , *Brain Research* § (2019).
- Tunbridge, E. M. (2010). The catechol-O-methyltransferase gene: Its regulation and polymorphisms. In *International Review of Neurobiology* (Vol. 95, pp. 7–27). Academic Press.
- Tunbridge, E. M., Farrell, S. M., Harrison, P. J., & Mackay, C. E. (2013). Catechol-O-methyltransferase (COMT) influences the connectivity of the prefrontal cortex at rest. *NeuroImage, 68*, 49–54.
- Turchi, J., Chang, C., Ye, F. Q., Russ, B. E., Yu, D. K., Cortes, C. R., ... Leopold, D. A. (2018). The Basal Forebrain Regulates Global Resting-State fMRI Fluctuations. *Neuron, 940-952.e4*.
- Turi, Z., Mittner, M., Opitz, A., Popkes, M., Paulus, W., & Antal, A. (2015). Transcranial direct current stimulation over the left prefrontal cortex increases randomness of choice in instrumental learning. *Cortex, 63*, 145–154.

- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 601–615.
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*, 14637.
- Vaiana, M., Goldberg, E. M., & Muldoon, S. F. (2019). Optimizing state change detection in functional temporal networks through dynamic community detection. *Journal of Complex Networks*, *7*, 529–553.
- van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016a). Pupil Diameter Tracks Lapses of Attention. *PLoS ONE*, *11*, e0165274.
- van den Brink, R. L., Nieuwenhuis, S., & Donner, T. H. (2018). Amplification and Suppression of Distinct Brainwide Activity Patterns by Catecholamines. *The Journal of Neuroscience*, *38*, 7476–7491.
- van den Brink, R. L., Pfeffer, T., Warren, C. M., Murphy, P. R., Tona, K.-D., van der Wee, N. J. A., ... Nieuwenhuis, S. (2016b). Catecholaminergic Neuromodulation Shapes Intrinsic MRI Functional Connectivity in the Human Brain. *The Journal of Neuroscience*, *36*, 7865–7876.
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 233121651983248.
- Vandenbergh, D. J., Persico, A. M., Hawkins, A. L., Griffin, C. A., Li, X., Jabs, E. W., & Uhl, G. R. (1992). Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics*, *14*, 1104–1106.
- Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological

- comparison in behaving monkeys. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*, 7866–7877.
- Vatansver, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default Mode Dynamics for Global Functional Integration. *Journal of Neuroscience*, *35*, 15254–15262.
- Verstynen, T. D., & Deshpande, V. (2011). Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. *NeuroImage*, *55*, 1633–1644.
- Wahn, B., Ferris, D. P., Hairston, W. D., König, P., Loewenfeld, I., Sirois, S., ... König, P. (2016). Pupil Sizes Scale with Attentional Load and Task Experience in a Multiple Object Tracking Task. *PLOS ONE*, *11*, e0168087.
- Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological Psychiatry*, *62*, 756–764.
- Waltz, J. A., Frank, M. J., Wiecki, T. V., & Gold, J. M. (2011). Altered probabilistic learning and response biases in schizophrenia: Behavioral evidence and neurocomputational modeling. *Neuropsychology*, *25*, 86–97.
- Warren, C. M., Eldar, E., van den Brink, R. L., Tona, K.-D., van der Wee, N. J., Giltay, E. J., ... Nieuwenhuis, S. (2016). Catecholamine-Mediated Increases in Gain Enhance the Precision of Cortical Representations. *The Journal of Neuroscience*, *36*, 5699–5708.
- Warren, C. M., van den Brink, R. L., Nieuwenhuis, S., & Bosch, J. A. (2017a).

- Norepinephrine transporter blocker atomoxetine increases salivary alpha amylase. *Psychoneuroendocrinology*, *78*, 233–236.
- Warren, C. M., Wilson, R. C., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., Cohen, J. D., & Nieuwenhuis, S. (2017b). The effect of atomoxetine on random and directed exploration in humans. *PLOS ONE*, *12*, e0176034.
- Weber, M. J., Messing, S. B., Rao, H., Detre, J. A., & Thompson-Schill, S. L. (2014). Prefrontal transcranial direct current stimulation alters activation and connectivity in cortical and subcortical reward systems: A tDCS-fMRI study. *Human Brain Mapping*, *35*, 3673–3686.
- Weinberger, D. R., Berman, K. F., & Daniel, D. G. (1992). Mesoprefrontal cortical dopaminergic activity and prefrontal hypofunction in schizophrenia. *Clinical Neuropharmacology*, *15*, 568A-569A.
- Weir, W. H., Emmons, S., Gibson, R., Taylor, D., & Mucha, P. J. (2017). Post-processing partitions to identify domains of modularity optimization. *Algorithms*, *10*.
<https://doi.org/10.3390/a10030093>
- Wells, H. G. (1895). *The Time Machine*. Project Gutenberg.
- Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 8456–8460.
- Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology. General*, *143*, 2074–2081.

- Witte, A. V., & Flöel, A. (2012). Effects of COMT polymorphisms on brain function and behavior in health and disease. *Brain Research Bulletin*, *88*, 418–428.
- Wolfensteller, U., & Ruge, H. (2012). Frontostriatal mechanisms in instruction-based learning as a hallmark of flexible goal-directed behavior. *Frontiers in Psychology*, *3*, 192.
- Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, *100*, 221–228.
- Wood, S. N. (2017). Generalized additive models: An introduction with R, second edition. In *Generalized Additive Models: An Introduction with R, Second Edition*. CRC Press.
- Yacubian, J., Sommer, T., Schroeder, K., Glascher, J., Kalisch, R., Leuenberger, B., ... Buchel, C. (2007). Gene-gene interaction associated with neural reward sensitivity. *Proceedings of the National Academy of Sciences*, *104*, 8125–8130.
- Yael, D., Vecht, J. J., & Bar-Gad, I. (2018). Filter-based phase shifts distort neuronal timing information. *ENeuro*, *5*. <https://doi.org/10.1523/ENEURO.0261-17.2018>
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–1165.
- Yu, A. J., & Cohen, J. D. (2009). Sequential effects : Superstition or rational behavior? *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, 1873–1880.
- Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*,

681–692.

Yue, Q., Martin, R. C., Fischer-Baum, S., Ramos-Nuñez, A. I., Ye, F., & Deem, M. W.

(2017). Brain modularity mediates the relation between task complexity and performance. *Journal of Cognitive Neuroscience*, 29, 1532–1546.

Yule, G. U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-

Series?--A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society*, 89, 1.

Záborszky, L., Gombkoto, P., Varsanyi, P., Gielow, M. R., Poe, G., Role, L. W., ...

Chiba, A. A. (2018). Specific Basal Forebrain–Cortical Cholinergic Circuits Coordinate Cognitive Operations. *The Journal of Neuroscience*, 38, 9446–9458.

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., & Herrmann, C. S. (2011).

Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: Combined behavioural and electrophysiological evidence. *BMC Neuroscience*, 12, 2.

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ...

Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68, vii–137.

Zerbi, V., Floriou-Servou, A., Markicevic, M., Vermeiren, Y., Sturman, O., Privitera, M.,

... Bohacek, J. (2019). Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron*, 0.

<https://doi.org/10.1016/j.neuron.2019.05.034>

Zhang, Z., Telesford, Q. K., Giusti, C., Lim, K. O., & Bassett, D. S. (2016). Choosing

Wavelet Methods, Filters, and Lengths for Functional Brain Network Construction.

PLOS ONE, 11, e0157243.

Zhao, M., Zhou, C., Chen, Y., Hu, B., & Wang, B.-H. (2010). Complexity versus modularity and heterogeneity in oscillatory networks: Combining segregation and integration in neural systems. *Physical Review E*, 82, 046225.

Zmigrod, S., Zmigrod, L., & Hommel, B. (2016). Transcranial direct current stimulation (tDCS) over the right dorsolateral prefrontal cortex affects stimulus conflict but not response conflict. *Neuroscience*, 322, 320–325.