NEURAL PROCESSES UNDERLYING THE FLEXIBLE CONTROL AND LEARNING OF ATTENTIONAL SELECTION

MARIANN OEMISCH

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

November 2017

© Mariann Oemisch, 2017

Abstract

In every-day life we are usually surrounded by a plethora of stimuli, of which only some may be relevant to us at a given moment in time. The dynamic interaction between internal factors, such as our previous experience and current goals, and external factors, such as salient sensory stimulation, determine where, how and what we attend to in our environment. This dissertation investigated some of the neural mechanisms that underlie successful goal-directed behavior in two conditions – 1. when attention was actively cued to a target stimulus, and 2. when the attentional target had to be actively and repeatedly learned, in macaque monkeys and in humans. In Chapter 2, I investigated inter-areal spiketrain correlations in neuron pairs across the fronto-cingulate cortex when macaque monkeys are cued to shift their attention to one of two target stimuli. I found that neuron pairs in anterior cingulate cortex (ACC) and dorsal prefrontal cortex (PFC) with similar spatial preferences correlate their spiketrains at the time when attention needs to be actively shifted, suggesting that the flexible interaction between these two areas may support successful covert attention shifts. In Chapter 3, I show that when the attentional target stimulus needs to be repeatedly learned and is defined by only one of several stimulus features, neurons in macaque frontal and striatal regions encode prediction error signals that carry specific information about the stimulus feature that was selected in the preceding choice. These signals may be involved in identifying those synapses that require updating to allow flexible adjustments in goal-directed behavior. In Chapter 4, I found that when humans must repeatedly learn the identity of an attentional target, a human event-related potential over visual cortex that is thought to index attentional target selection, selectively decreases after successful learning, in particular for the distracting stimulus, and selectively increases for the target stimulus following negative feedback during learning. Overall, this dissertation provides novel insights into some of the complex neural mechanisms that support flexible control and learning of attention across brain regions of the human and non-human primate brain.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my Ph.D. supervisor Dr. Thilo Womelsdorf for his relentless support, guidance, and patience throughout the last five years. Thank you for sharing your invaluable knowledge and expertise; it has helped me grow immensely as a scientist and as an individual.

I would like to thank my current and previous supervisory committee members Dr. Kari Hoffman, Dr. Douglas Crawford, and Dr. Keith Schneider for their valuable and much appreciated advice and support. I would also like to thank Dr. Anna Schubö, who was so generous to host me in her lab for my IRTG project in Marburg, Germany.

I am grateful to all the previous and current members of the Womelsdorf lab for their endless scientific and personal support, and for the much-enjoyed distractions when needed. I would like to especially thank Dr. Stephanie Westendorff for her excellent training and guidance throughout my first three years in the lab, Benjamin Voloh for the inspiring scientific and non-scientific discussions in and outside of the lab, Dr. Marcus Watson for his scientific and personal guidance in the last two years, and finally Samira Azimi and Ali Hassani for the invaluable and much appreciated help with animal training and data collection in the last weeks and months – I could not have done this without you! I would also like to thank the entire animal care and veterinary staff for their outstanding expertise and help with the animal care.

Last but not least, I would like to extend my warmest thanks to my friends and family, and especially to my partner Marc, for the never-ending patience and much needed support throughout the last weeks, months, and years.

Table of Contents

Abstract ii
Acknowledgements iii
List of Figures vii

Chapter 1 General Introduction 1
1.1 Mechanisms of attention 1
1.1.1 Goal-directed attention 1
1.1.2 Experience-driven attention 3
1.2 Principles of value learning 6
1.2.1 Reinforcement learning 7
1.2.2 Theories of attentional learning 9
1.3 Brain networks underlying attention mechanisms 12
1.3.1 Brain regions that support 'attention for action' system 12
1.3.2 Brain regions that support 'attention for learning' system 14
1.3.3 Brain regions that support 'attention for liking' system 16
1.3.4 Anatomical connectivity of brain regions involved in value-based attention 17
1.4 Scope of this dissertation 18
1.5 References 21

Chapter 2 Inter-areal spiketrain correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts 30
2.1 Summary 31
2.2 Introduction 31
2.3 Methods 33
2.3.1 Procedures 33
2.3.2 Behavioral paradigm 34
2.3.3 Anatomical reconstruction 36
2.3.4 Data analysis 37
2.4 Results 46
2.4.1 Behavioral performance 47
2.4.2 Spiketrain correlations increase when selective attention is deployed 47
2.4.3 Temporal specificity of spiketrain correlations 50

2.4.4 Anatomical specificity of spiketrain correlations 51
2.4.5 Correlated firing is maximal during attention shift....51
2.4.6 Asymmetry of spiketrain correlations....52
2.4.7 Spiketrain correlations of cell pairs with similar spatial attention preferences 54
2.4.8 Cell-type specificity of spiketrain correlations 56
2.4.9 Spiketrain correlations are not related to expected reward magnitude 58
2.4.10 Spiketrain correlations show no apparent relation to the occurrence of burst firing 59
2.4.11 Spiketrain correlations show no apparent relation to LFP theta-gamma phase amplitude correlations....59
2.5 Discussion 60
2.6 References....66

Chapter 3 Signed and unsigned prediction error signals in macaque fronto-striatal circuit carry information about the preceding feature choice....70
3.1 Summary 71
3.2 Introduction....72
3.3 Methods....74
3.3.1 Electrophysiological recordings....74
3.3.2 Behavioral paradigm 75
3.3.3 Stimuli....77
3.3.4 Data analysis 77
3.4 Results....88
3.4.1 Behavioral performance....88
3.4.2 Task variables encoded in fronto-striatal areas....89
3.4.3 Multiplexing of task information 92
3.4.4 RL model fit....95
3.4.5 Non-specific prediction errors in fronto-striatal areas 95
3.4.6 Feature-specific prediction errors in fronto-striatal areas 99
3.4.7 Area-dependent time course of prediction error signals 105
3.4.8 Time-course of trial outcome encoding versus prediction error encoding 111
3.4.9 Cell-type specificity of prediction error signals....114
3.5 Discussion 118
3.6 References....126

Chapter 4 Changes of attention during value-based reversal learning are tracked by N2pc and feedback-related negativity 132
4.1 Summary 133
4.2 Introduction 133
4.3 Methods 138
4.3.1 Participants 138
4.3.2 Apparatus and stimuli 138
4.3.3 Procedure 141
4.3.4 EEG recording 142
4.3.5 Data analysis 143
4.4 Results 146
4.4.1 Reversal learning 146
4.4.2 Attention deployment changes with learning 147
4.4.3 Low value feedback is followed by increased attentional target selection 151
4.4.4 Feedback processing is increased *during* learning 153
4.5 Discussion 156
4.6 References 163

Chapter 5 General Discussion 168
5.1 Summary 168
5.2 In light of attention mechanisms 169
5.3 In light of uncertainty 170
5.4 Neural network of attention and learning 172
5.5 Interpretational limitations 173
5.6 Future directions 174
5.7 Implications for understanding brain health 175
5.8 Concluding remarks 176
5.9 References 178
Appendix A Supplementary Figures Chapter 3 181
Appendix B Supplementary Methods and Results Chapter 3 192
Appendix C Co-Authorship 196

List of Figures

Figure 2.1 Task, Recording Locations and Example JPSTHs. 35
Figure 2.2 Average Firing Correlations with JPSTH and Pearson correlation. 39
Figure 2.3 Firing Correlations and their temporal and spatial specificity. 49
Figure 2.4 Asymmetry of Correlated Firing. 53
Figure 2.5 Firing Correlations carry spatial attention information. 55
Figure 2.6 Cell type specificity of firing correlations in ACC/PFC. 57
Figure 2.7 Sources of correlated firing of cells pairs from different cortical fields in ACC and dorsal PFC. 65
Figure 3.1 Behavioral task, performance and recording locations. 76
Figure 3.2 Task variables encoded at time of outcome. 90
Figure 3.3 Second order encoding of task variables. 93
Figure 3.4 Reinforcement learning model. 96
Figure 3.5 Non-specific prediction error encoding. 97
Figure 3.6 Examples of neurons encoding feature-specific PE signals. 100
Figure 3.7 Feature-specific prediction error encoding. 102
Figure 3.8 Summary of PE time course encoding. 107
Figure 3.9 Time courses of feature-specific PE signals. 110
Figure 3.10 Latency comparison of PE- and reward-encoding. 112
Figure 3.11 Cell-types encoding prediction errors in *monkey H*. 115
Figure 3.12 Cell-types encoding prediction errors in *monkey K*. 117
Figure 4.1 Hypotheses for learning-related changes in behavior, N2pc, and FRN amplitudes. 137
Figure 4.2 Value-based reversal learning task and example block performance computed with EM algorithm. 139
Figure 4.3 Behavioral performance of the reversal learning task. 148
Figure 4.4 Lateralized N2pc components. 150
Figure 4.5 Effect of previous trial feedback on N2pc amplitude. 152
Figure 4.6 FRN amplitudes. 155

Chapter 1

General Introduction

1.1 Mechanisms of attention

Considering the mass of information that continuously surrounds us in the world, attention provides the gateway that allows us to select a small subset of that information for action, further analysis and learning (Mitchell and Le Pelley, 2010). In everyday life, the focus of our attention is guided by internal factors, such as our experience and our current goals, and by external factors such as salient sensory stimulation in our environment. The dynamic interaction of these factors determines where, how and to what we attend to in our environment (Corbetta and Shulman, 2002). For example, the goal of finding the red marker in our apartment may guide our attention to specific features and spatial locations. All items of the color red may become selectively attended to, and our cluttered desk may become the main focus of our spatial attention because it is the most likely location for finding the red marker. In addition, based on our recent experience, the turquoise mug on our desk that has been a valuable coffee suspender over the last weeks may draw some of our attention even though it does not match our current goals (Awh et al., 2012). And finally, the precariously stacked dishes in the dish rack that slowly started to slip and finally crashed onto the kitchen counter will take immediate precedence over our current goals and experience and draw our attention instantly to the cause of the sudden noise. These different factors driving attention can be dissociated into goal-directed attention, experience-driven attention, and saliency-driven attention. I will now introduce the concepts of the former two, as these are most relevant to the thesis at hand.

1.1.1 Goal-directed attention

It is known that observers are better at detecting objects when they have information about the object's feature, such as its location, motion or color, in advance (Eriksen and Hoffman, 1973; Ball and Sekuler,

1980; Posner, 1980). Posner (1980) for instance showed that covert shifts of attention to an empty space led to faster processing of stimuli (reduced reaction time) at the attended location versus at an unattended location. Since then, the brain networks that may underlie this flexible allocation of attention during goal-directed behavior have been at the center of investigation in a plethora of experimental studies. It seems that attention affects sensory processing at all levels of the visual hierarchy, from relatively low-level areas, such as the thalamic nuclei to higher level cortical areas (e.g. Desimone and Duncan, 1995; Kastner and Ungerleider, 2000; Reynolds and Heeger, 2009; Miller and Buschman, 2012; Saalmann et al., 2012). It has for instance been shown that neural responses to a stimulus inside a neurons receptive field are enhanced when attention is directed inside the receptive field compared with outside. This has been demonstrated for neurons in visual cortex areas V1 (Motter, 1993), V2 (Motter, 1993; Luck et al., 1997), and V4 (Connor et al., 1996; Luck et al., 1997; McAdams and Maunsell, 1999), as well as in downstream areas such as area MT (Treue and Maunsell, 1996; Treue and Martínez Trujillo, 1999) and lateral intraparietal area (LIP) (Colby et al., 1996). Attention can bias neural activity not only in favor of an attended spatial location, but also in favor of an attended feature. Neurons in V4 for instance selectively increased their firing rates for responses to stimuli that matched the attended feature (e.g. color or luminance) compared with stimuli that did not match the attended feature (Motter, 1994). Similar effects have been shown for additional features such as orientation (Haenny and Schiller, 1988) and motion direction (Ferrera et al., 1994). These results suggest that goal-directed attention mechanisms operate by selectively enhancing neural processing of attended stimulus features. It has since been shown that attention may also operate via a suppression of unwanted information, an increase in baseline activity at attended locations, and an increase in response sensitivity (Kastner and Ungerleider, 2000; Reynolds and Heeger, 2009). Additionally, it has been shown that attentional modulation does not solely include a modulation of firing rate, but also an increase of the reliability of firing of individual neurons (Mitchell et al., 2007), and a decrease of co-variability across pairs of neurons (Cohen and Maunsell, 2009; Mitchell et al., 2009). Importantly, it has been suggested that modulation by goal-directed attention originates in the

frontal cortices and is relayed via the parietal cortex to earlier visual areas (Corbetta and Shulman, 2002; Miller and Buschman, 2012). It has for instance been shown that neurons in lateral prefrontal cortex (LPFC) and the frontal eye field (FEF) signal shifts of goal-directed attention with shorter latency than neurons in LIP (Buschman and Miller, 2007). Complimentary to that, human frontal cortex generates a P300 event-related potential (ERP), an ERP previously linked to the voluntary detection of task-relevant stimuli, with peak amplitude and shorter latency compared with human parietal cortex, during goal-directed attention shifts (Li et al., 2010). Further downstream, attentional modulation occurs earlier in LIP neurons compared with MT neurons (Saalmann et al., 2007), and earlier in visual area V4 compared with V2, and earlier in V2 compared with V1 (Buffalo et al., 2010), when monkeys voluntarily shift attention. During goal-directed shifts of attention, frontal cortical regions therefore seem to be involved in the decision process of what and where attention will be shifted to. Within frontal regions, the integration of information about the location and value of attentional targets has been found at the intersection of LPFC, ventromedial PFC (vmPFC) and anterior cingulate cortex (ACC) (Kaping et al., 2011). ACC in particular has been suggested to contribute to control signals that monitor and guide attentional selection (Buckley et al., 2009; Alexander and Brown, 2011; Shenhav et al., 2013). This distributed network of brain regions involved in goal-directed attention shifts may communicate through synchronous activity at the level of neural ensembles through rhythmic oscillations (Fries et al., 2001; Womelsdorf et al., 2006; Saalmann et al., 2007; Gregoriou et al., 2009; Voloh et al., 2015; Womelsdorf and Everling, 2015) and potentially at the level of single neurons through inter-areal spiketrain correlations (**Chapter 2**, Oemisch et al., 2015).

1.1.2 Experience-driven attention

Less studied than goal-directed attention, is what can be described as experience-driven attention, in which selection history and reward learning can affect attentional prioritization, sometimes independent of current goals (Chun and Turk-Browne, 2007; Awh et al., 2012). One example of recent selection

history's influence on attention is inter-trial priming, which refers to the observation that stimuli or features that have recently been selected for attentional processing, are more efficiently attended in the following trials. Maljkovic and Nakayama (1994, 1996) found that when participants searched for a target defined by a feature (e.g. color or spatial frequency) among distractors, search for the same feature was more efficient in subsequent trials (i.e. reduced reaction time and enhanced accuracy). Importantly, this priming effect was robust and even strongest when search features predictably alternated every trial, suggesting that this priming was not under voluntary control. Inter-trial priming is thought to be mediated by implicit memory traces (Schacter and Buckner, 1998). Neurophysiological evidence in humans (Eimer et al., 2010) and monkeys (Bichot and Schall, 1999) suggests that visual processing of the target as well as inhibition of distractors occurs sooner and is more pronounced under circumstances of priming. Neurons in FEF discriminated better and earlier between target and distractors in trials in which target features were repeated (Bichot and Schall, 1999, 2002). However, in addition to the FEF, it seems that the network typically associated with attentional control is selectively suppressed during trials in which target features are repeated, including the ACC and intraparietal sulcus (Kristjansson et al., 2007), a somewhat counter-intuitive neural signature of priming that is often referred to as repetition suppression (see for review Kristjansson and Campana, 2010). A possible explanation for this phenomenon is that voluntary shifts of attention, which are required in trials in which the target feature switches, may require a reset of perceptual systems to shift them into a new 'attractor' state, causing increased and delayed activity (Serences and Yantis, 2006).

In contrast to inter-trial priming, the phenomenon of inhibition of return (IOR) refers to the observation that, following an initial capture of attention, subjects become less efficient at detecting and discriminating targets at locations that were cued more than approximately 250ms before (Posner and Cohen, 1984; Posner et al., 1985). Thought to drive orienting to novel locations, the detailed mechanisms underlying IOR are not yet fully understood (Raymond M. Klein, 2000; Lupiáñez et al., 2006), but the

superior colliculus (SC) seems to play a prominent role in its mediation (Sapir et al., 1999; Dorris et al., 2002; Fecteau et al., 2004).

As well as very recent selection history, implicit and explicit memories of previously experienced arrays or scenes that extend further into the past can influence attention. The contextual cuing effect, for example, refers to the finding that subjects are faster at detecting visual targets among distractors, when these appear repeatedly in the same configuration, even without being aware of the regularities of these configurations, and presumably through implicit memory (Chun and Jiang, 1998; Chun, 2000). Similarly, explicitly learned memories of targets in complex visual scenes speed up repeated attentional target selection more than cued target selection (Summerfield et al., 2006). This type of memory-guided attention and contextual cuing seem to be mediated by the medial temporal lobe and fronto-parietal network (Chun and Phelps, 1999; Summerfield et al., 2006). Moreover, there is substantial literature showing the ways in which overt attention (i.e. oculomotor behavior) is changed by experience, generally over much longer time scales (see for review Tatler et al., 2011; Vickers, 2016). For example, elite athletes tend to fixate critical objects or locations earlier and for longer than lower skilled performers (Vickers, 2007). While the previous examples of how experience affects attentional selection are important, they are of less immediate relevance to my thesis and for that reason by no means introduced comprehensively.

As mentioned at the beginning of this section, attentional prioritization can also be powerfully affected by previously learned reward values. In fact, there seems to be a pronounced selection bias for previously rewarded features, stimuli, or actions (e.g. Dayan and Balleine, 2002; Shuler and Bear, 2006; Seitz et al., 2009; Navalpakkam et al., 2010). Kaping et al. (2011) showed that performance accuracy increased when covert attention was shifted to a high-value as opposed to a low-value reward target, and neural activity in vmPFC and ACC was modulated by the value of the attentional target. This intuitively makes sense from the stand point of incentive motivation (Pessoa and Engelmann, 2010), whereby reward enhances perceptual and executive control processes that lead to more efficient goal-directed behavior. Reward in

this case provides the motivational significance that then leads to voluntary attentional orienting towards those rewarded stimuli (Awh et al., 2012). However, it has recently been shown that reward can also affect subsequent selection that is not in line with the voluntary selection goal of the observer (Hickey et al., 2010; Anderson et al., 2011). Anderson and colleagues (2011) for instance showed that a distractor, whose feature was previously associated with a high reward, slowed target processing more and was therefore more distracting than a distractor whose feature was previously associated with a low reward. Likewise, target processing was increased (e.g. reduced reaction time) following a high reward outcome compared with a low reward outcome, when the target feature remained the same, but was reduced following a high reward outcome compared with a low reward outcome when the target feature changed (Hickey et al., 2010). This suggests that reward value per se has a pronounced impact on attentional selection, and the two may indeed be closely intertwined concepts (Maunsell, 2004). It seems reasonable to assume that a critical goal of attentional prioritization is reward maximization and loss minimization (Navalpakkam et al., 2010). In a naturalistic setting, where the values of stimuli and actions may change frequently, it seems therefore pivotal to continuously update the choice criteria with regards to the stimuli we attend to, in order to reach that goal of reward maximization. The question then becomes how we know and learn what is most valuable to attend to in a changing and volatile environment (Gottlieb, 2012).

1.2 Principles of value learning

The theory of learning about one's environment and in particular about the values of stimuli and actions is routed in the framework of classical and instrumental conditioning (Dayan et al., 2000). Classical conditioning experiments aim to understand the processes by which an animal learns the predictive relationship between a stimulus and its consequences. Traditionally these experiments pair the occurrence of a conditioned stimulus (CS, e.g. the flash of a light) with that of an unconditioned stimulus linked to

primary reward (US, e.g. sucrose), such that over time the animal comes to reliably predict the US when experiencing the CS (Dickinson, 1980; Mackintosh, 1983), presumably leading to a shift of selective attention to the CS upon appearance (Mackintosh, 1975; Dayan et al., 2000). Instrumental conditioning experiments, in contrast to classical conditioning experiments, investigate how animals learn the predictive relationship between an action and an outcome (US). These experiments require an explicit action by the animal, such as the pressing of a lever in order to receive a sugar pellet (Balleine et al., 2009). The action the animal performs then becomes a conditioned response. In conditioning experiments in general therefore the animal learns that a stimulus or an action predicts a specific outcome, and learns to adjust its behavior accordingly. A computational account that describes learning about the environment, via forming of such associations between stimuli and outcomes and actions and outcomes, is reinforcement learning.

1.2.1 Reinforcement learning

Reinforcement learning (RL) is the process of repeatedly performing actions on the environment, and updating expectations and values of actions based on the feedback that is received after performing those actions. More formally, RL is the adaptive process of an agent (e.g. an animal) constantly estimating the values of its options based on the repeated experience with these options to maximize the outcome of future choices. The computational framework of RL that originated in machine learning (Sutton and Barto, 1998), has been particularly successful in describing natural (and artificial) decision-making systems (Dayan and Daw, 2008; Dayan and Niv, 2008). The core idea of RL is captured in the original Rescorla-Wagner equation, which describes the updating of a value function at time t (V_t) based on previous estimates of this value function (V_{t-1}) and a learning term ($\beta*\delta$) that includes learning rate (β) and a prediction error (δ) (Rescorla and Wagner, 1972):

$$V_t = V_{t-1} + \beta * \delta$$

The value function refers to the empirical estimate of the sum of future rewards (Lee et al., 2012). A given action or choice is then made based on its value function; the greater the associated value function about a particular action, the more likely it is that the action is performed (for instance the lever press that reliably leads to a sucrose reward). Value functions are updated based on the outcome received from performing that particular action or choice (Lee et al., 2012). The learning term above ($\beta*\delta$) is comprised of the learning rate (β) and a prediction error term (δ). The learning rate determines how quickly an agent updates its predictions and depends on task properties such as volatility. The prediction error term reflects the discrepancy between the predicted and the actual outcome received from the most recent choice. The prediction error then provides a trigger for learning (updating of expectations, value functions) so that future estimates are more accurate (Niv, 2009). In a now classic set of experiments, a neural correlate of prediction errors has been found in dopamine neurons in the ventral tegmental area (VTA) of the midbrain (Schultz et al., 1993; Fiorillo et al., 2003; Tobler et al., 2005). These dopamine neurons fire action potentials whenever an unexpected reward is delivered, a response that declines in magnitude with increasing predictability of reward, and finally ceases once reward delivery is fully predicted (Fiorillo et al., 2003; Bayer and Glimcher, 2005). Instead the dopamine neuron's response shifts forward in time to the occurrence of the stimulus that predicts reward. In this case, if, at the time of the predicted reward occurrence, the reward is larger than expected, the neurons increase their firing and signal a positive prediction error. If the reward delivered is smaller than expected, neurons decrease their firing and signal a negative prediction error (Fiorillo et al., 2003; Nomoto et al., 2010). Since then, prediction error-like signals have been observed across many different brain regions (e.g. Matsumoto et al., 2007; Hare et al., 2008; Asaad and Eskandar, 2011; Hayden et al., 2011; Kennerley et al., 2011), and potentially represent a common computation of the brain that help us make sense of the world (Friston, 2005; Den Ouden et al., 2012). The link back to attention comes from those studies that are concerned with assessing the relevance of stimuli for predicting or failing to predict outcomes, such that those stimuli that are predictively useful become selectively attended to (Mackintosh, 1975; Pearce and Hall, 1980).

1.2.2 Theories of attentional learning

Two classic theories that aim to describe how selective attention towards stimuli changes with associative learning were proposed by Mackintosh (1975) and by Pearce and Hall (1980). Both theories hold that attention is modulated by the predictive validity or predictive significance of stimuli, however, the nature of the modulation proposed by those initial theories is opposing. According to Mackintosh (1975), attention towards a stimulus that is better correlated with reinforcement and therefore a better predictor of reinforcement than other stimuli will increase with learning, while attention towards a stimulus that is a poor predictor of reinforcement will decrease with learning. The Mackintosh theory is based on a set of experiments that required animals to learn discrimination problems (Mackintosh and Little, 1969; Dias et al., 1996; George and Pearce, 1999; Dopson et al., 2010). In the original Mackintosh and Little (1969) study for instance, animals had to discriminate compound stimuli that were characterized by multiple features (e.g. color and shape) whereby only one feature was linked to reward at a given time (e.g. blue rewarded). When the feature-reward association changed, animals were quicker to learn the new rule if the rewarded dimension remained the same (e.g. yellow rewarded; intra-dimensional shift), compared to when the rewarded dimension changed (e.g. squared shape rewarded; extra-dimensional shift). This suggested that animals that were trained on a discrimination problem paid more attention to those stimuli that were relevant to its solution, and led Mackintosh (1975) to suggest that greater predictability led to greater attention. Contrary to this theory, the Pearce and Hall (1980) theory posits that stimuli that accurately predict future events should receive little attention, while those stimuli that are inaccurate predictors of future events should receive more attention to govern learning. This theory is grounded in experiments that showed that animals that were conditioned with a partial reinforcer (e.g. 50% probability of being followed by reward) learned faster about those reinforcers afterwards, compared with animals that had been conditioned with a continuous reinforcer (e.g. 100% probability of being followed by reward) (Hall and Pearce, 1979; Pearce and Hall, 1979; Wilson et al., 1992; Hogarth et al., 2008; Haselgrove et al., 2010). Rats in the Haselgrove study, for instance, learned a discrimination problem

between two stimuli faster if those stimuli were previously conditioned with partial reinforcement than when they were conditioned with continuous reinforcement (Haselgrove et al., 2010). This suggested that attention to stimuli whose consequences are fully predicted is reduced, while attention to stimuli whose consequences are poorly predicted remains high. On a trial-by-trial basis, this suggests that attention should be increased following unexpected and surprising events that led to prediction errors in order to facilitate learning and reduce future errors in prediction (Pearce and Hall, 1980). Indeed, in humans, a trial-dependent change in the allocation of attention following uncertain outcomes has been observed over the visual cortex when subjects have to identify features of visual stimuli (**Chapter 4,** Oemisch et al., in press). Both theories have received substantial support throughout the years and efforts have been made to reconcile their findings and implications (Dayan et al., 2000; Le Pelley, 2004; Pearce and Mackintosh, 2010). A popular framework that brings the findings from both theories together suggests that, dependent on the circumstances, different attention mechanisms may be implemented (Dayan et al., 2000; Pearce and Mackintosh, 2010; Gottlieb, 2012). One of these attention mechanisms is then more concerned with prediction and action, while the other is more concerned with uncertainty and learning (Maddux et al., 2007; Hogarth et al., 2010). In fact, a third attention mechanism has been suggested that is concerned with 'liking' and is independent of the two other mechanisms (Holland and Gallagher, 1999; Vuilleumier, 2005; Hogarth et al., 2010; Gottlieb, 2012). The idea is that the attention mechanism for 'action' is driven by a more 'automatic' form of attentional selection, that allows efficient detection of predictive stimuli and rapid action selection, for instance when animals are confronted with competitive combinations of multiple stimuli. In this case, reliable predictors should receive the most attention to determine a fast response, and this relationship strengthens the more reliable the predictor is (Dayan et al., 2000; Hogarth et al., 2010). On the other hand, following failures in prediction, unreliable stimuli may command a more 'controlled' form of attentional selection, which drives increased attention to the most inaccurate predictors to facilitate learning about the predictive environment and reduce uncertainty (Dayan et al.,

2000; Hogarth et al., 2010; **Chapter 4**). Pearce and Mackintosh (2010) proposed an additional associability parameter α implemented in the Rescorla-Wagner equation:

$$V_t = V_{t-1} + \boldsymbol{\alpha} * \beta * \delta$$

This associability parameter reflects the stimulus-specific learning rate. While the learning rate β applies globally to a task or context, α is a property of an individual cue or stimulus. During periods when the 'attention for action' mechanism is in place, associability increases with reliability, while when the 'attention for learning' mechanisms operates, associability increases with uncertainty (Pearce and Mackintosh, 2010; Gottlieb, 2012).

Both, the 'attention for action' and the 'attention for learning' mechanisms described above assign value to stimuli based on their relevance for future actions and fall under the category of goal-directed attention described at the beginning of this chapter. Importantly, these mechanisms are not directly dependent on expected reward. The third proposed mechanism of attention, briefly mentioned above, is the 'attention for liking' mechanism, which assigns attentional priority to those stimuli that are associated with reward and has traditionally been described in the field of emotion (Vuilleumier et al., 2003; Engelmann and Pessoa, 2007; Brosch et al., 2008). This concept describes the findings that stimuli with increased biological value command increased attention, such as stimuli associated with larger quantities of food (Mogg et al., 1998; Morris and Bouton, 2006) or drugs of abuse (Field et al., 2004). Similarly, aversive stimuli or cues can be followed by attentional avoidance (Pflugshaupt et al., 2005; Weierich et al., 2008). Peck et al. (2009) had monkeys perform a cued choice task, in which two placeholders were presented to the left and right of a central fixation point; the animals had to saccade to, and at the time when, one of the placeholders was illuminated. Importantly, beforehand a response-independent cue located at either of the two placeholder locations indicated the outcome of the upcoming choice (rewarded or not rewarded). Peck and colleagues found that choices towards the illuminated stimulus were fastest and most accurate when the outcome cue occurred at its location and indicated reward. In contrast, animals were slowest and

least accurate responding towards a stimulus at the same location as a previous cue that indicated no reward delivery, suggesting attention drawn away from that location in an avoidance-like behavior. The proposed 'attention for liking' mechanism falls under the category of experience-driven attention that was described earlier in this chapter.

Overall, it seems that there are several mechanisms, that likely operate in tandem, by which the brain decides what and where to attend to (Leong et al., 2017). In light of the vast range of complex and flexible behavior possible, it intuitively makes sense that we depend on multiple accounts of assigning value to stimuli, actions or contexts, depending on the current behavioral situation.

1.3 Brain networks underlying attention mechanisms

We have already discussed some brain regions that seem to be involved in directing attention earlier in this chapter. We have for instance learned about evidence that goal-directed attention is initiated in the frontal regions and relayed to the parietal cortex and then to visual areas (Corbetta and Shulman, 2002; Miller and Buschman, 2012). But whether the purpose of attention is to promote target selection in a familiar task ('action'), to promote gathering of new information about uncertain predictors ('learning'), or to simply obtain reward ('liking') has rarely been distinguished at the neural level.

1.3.1 Brain regions that support 'attention for action' system

When the purpose is to perform a correct action or target selection, it seems attention is allocated to the most relevant and reliable information. It is however not well understood how and where the brain encodes cognitive reliability and relevance. Evidence for the encoding of reliability has been found in several brain regions (Bromberg-Martin and Hikosaka, 2009; Blanchard et al., 2015; Foley et al., 2017). Bromberg-Martin and Hikosaka (2009) for instance found that many primate dopamine neurons that

encode outcome also encode the reliability of a cue in predicting the outcome. In this task, animals were presented with an informative cue that predicted the outcome of the trial with a 100% probability in one set of trials, and with an uninformative cue that predicted the outcome of the trial with a 50% probability in another set of trials. Dopamine neurons in the VTA encoded the reliability or 'informativeness' of the cue before information about the outcome was revealed, suggesting dopamine neurons as a potential substrate for encoding reliability of cues, at least when those cues predict primary rewards (Bromberg-Martin and Hikosaka, 2009). A similar study revealed that neurons in orbital frontal cortex (OFC) are also selective for the reliability of cues in predicting primary reward outcomes (Blanchard et al., 2015). Unlike in dopamine neurons however, separate populations of OFC neurons encoded reliability and outcome value. Finally, a recent study showed that neurons in LIP also encode the reliability of cues, however not linked directly to primary reward, but instead to the reliability of providing relevant information at a later point in time (Foley et al., 2017).

The neural mechanisms for encoding of relevance are not well understood, but research generally points towards the frontal cortex. Activity in ACC for instance has been shown to distinguish between cues that predict the target location of an upcoming saccade well, and are therefore relevant to the preparation of an action, and those cues that do not (O'Reilly et al., 2013). Evidence also comes from studies that investigate the problem of credit assignment (Gottlieb, 2012). Credit assignment describes the computation of correctly assigning credit of an outcome to the source of that outcome in the previous choice, or to the relevance of a source in causing a given outcome (Walton et al., 2011), and has been suggested to be computed in regions of the frontal cortex, most notably in OFC (Rushworth et al., 2011; Chau et al., 2015). Furthermore, animals with damage to the prefrontal cortex are impaired in attending to relevant stimuli (Rossi et al., 2009) or task rules (Buckley et al., 2009), especially when these switch repeatedly. Frontal brain regions, in addition to parietal and subcortical regions, may therefore be primary drivers for attentional orienting to relevant and reliable stimuli or actions, when the current purpose of attention is to drive target selection in a familiar task environment (see also **Chapter 2**).

1.3.2 Brain regions that support 'attention for learning' system

During learning, it seems that attention is drawn towards the more uncertain stimuli or actions, likely to promote gathering evidence about their nature in predicting reinforcing events. Neural responses that seem to encode uncertainty or surprise have been observed in many different brain regions. As mentioned earlier on, dopamine neurons are well known for encoding signed prediction error signals (PEs) (e.g. Fiorillo et al., 2003). However, prediction error-like signals have been observed ubiquitously across many brain regions in relation to processing of sensory signals, motor actions, value and cognitive control, and can be differentiated into signed quantities that carry information about the valence of the outcome, and unsigned quantities that signal surprise independent of valence (den Ouden et al., 2012). Various sensory cortical areas for instance signal surprise about unexpected sensory events (Summerfield and Koechlin, 2008; Wacongne et al., 2011; Kok et al., 2012). Higher cortical areas, such as the lateral PFC, ACC and OFC have been shown to encode signed reward PEs and surprise about choice outcomes (Hayden et al., 2011; Kennerley et al., 2011), signed PEs about action values (Matsumoto et al., 2007; Sul et al., 2010), surprise about abstract decision states (Gläscher et al., 2010; Daw et al., 2011), surprise about abstract rule violations (Fletcher et al., 2001; Corlett et al., 2004), and surprise and signed reward PEs about feature values (**Chapter 3**). Subcortical areas, such as the ventral striatum in particular, but also caudate nucleus, amygdala, VTA and lateral habenula, have been shown to encode choice outcome signed reward PEs and surprise (Yacubian, 2006; Hare et al., 2008; Apicella et al., 2009; Asaad and Eskandar, 2011), signed punishment PEs (Matsumoto and Hikosaka, 2007), signed PEs about abstract decision states (Daw et al., 2011), signed stimulus PEs (Guo et al., 2016), signed fictive PEs (Lohrenz et al., 2007), signed PEs about action values (Stalnaker et al., 2012), and surprise and signed reward PEs about feature values (**Chapter 3**). The previous literature suggests that surprise and prediction error signals occur wide-spread across the brain. An increase in attention towards uncertain stimuli or actions following uncertain outcomes may therefore be driven by those brain regions that reflect the source of uncertainty that was experienced.

A direct attempt at dissociating the allocation of 'attention for learning' and 'attention for action' in the same animals has been made in behavioral lesion studies of the rat (Maddux et al., 2007; Maddux and Holland, 2011). In these studies, rats were probed in two stages; in an initial 5-choice serial reaction time task (5-CSRT), rats were presented with five port lights and corresponding response ports. Following a brief illumination of one of the five port lights required the animal to nose-poke the corresponding response port to receive food pellets. Importantly, some port lights were consistently reinforced while others were only partially reinforced (50%). Increased accuracy and reduced reaction time when the consistently reinforced ports were illuminated (in sham rats), indicated that when an immediate action was required, the animal attended more to the relevant stimuli that were good predictors of reinforcement (Mackintosh, 1975; Dayan et al., 2000). The second stage probed attention for new learning indirectly. Rats were now conditioned to associate two tones that were presented concurrently with the partially and consistently reinforced light ports, respectively, with food delivery (no required action). According to Pearce and Hall (1980), the partially reinforced light port should retain more attention, and therefore the ability to learn about its association, in which case it should overshadow conditioning to its tone-partner, whereby overshadowing refers to the phenomenon that an equally relevant but more salient stimulus can decrease or completely prevent conditioning of a less salient stimulus (Pavlov, 1927; Mackintosh, 1976). Indeed, sham rats were worse at associating the light that was presented concurrently with the partially reinforced light port with food, than associating the tone that was concurrently presented with the consistently reinforced light port with food. Maddux and colleagues (2007) lesioned rats in the medial prefrontal cortex (mPFC), central amygdala, or posterior parietal cortex (PPC) and assessed their behavior in the aforementioned task. They found that rats with a lesioned mPFC and central amygdala responded equally to partial and consistent reinforcement, suggesting that they were impaired at attending preferentially to the more relevant stimulus, which was not the case for rats with a lesioned PPC. On the contrary, rats with lesioned central amygdala and PPC showed no overshadowing in the second stage,

suggesting that they did not attend more to the partial reinforcer, which was not the case for rats with lesioned mPFC. This study suggests a specific involvement of the mPFC in allocating attention during the control of action and of the PPC in allocating attention during new learning, while the amygdala seems to be involved in both processes (Maddux et al., 2007). A follow-up study suggested a dissociation within the mPFC, whereby the ventral mPFC seems to be involved in the allocation of attention during control of action, while the dorsal mPFC seems to be involved in the allocation of attention during new learning (Maddux and Holland, 2011). It is unclear whether such a distinction holds in the primate brain and how it is implemented at the neuron level.

1.3.3 Brain regions that support 'attention for liking' system

The third mechanism to drive attention was referred to as 'attention for liking', and described the inherent attraction of attention towards stimuli of emotional or conditioned association (Vuilleumier, 2005; Della Libera and Chelazzi, 2009; Hickey et al., 2010). Evidence that the parietal cortex is involved in this type of attention mechanism comes from the study by Peck and colleagues (2009) that was mentioned a little earlier already. Here monkeys showed a behavioral bias towards spatial locations that had been cued with a positive outcome cue, and away from spatial locations that had been cued with a negative outcome cue. At the same time, neurons in LIP showed a transient visual response that was higher for the positive outcome cue than for the negative outcome cue. In addition, LIP neurons showed sustained activity after the outcome cue disappeared, which was largest at the location that was cued with a positive outcome cue and was smallest at the location that was cued with a negative outcome cue. This suggested that LIP neurons reflected the increased attentional allocation to seemingly higher valued spatial locations and away from seemingly lower valued spatial locations, even if these attention shifts were irrelevant or even detrimental to task performance (Peck et al., 2009). In an economic decision making task, attention elicited by a visual cue, as measured using an EEG component above human visual cortex, was correlated with the predictive association of that cue with positive outcome (San Martín et al., 2016). This could

potentially suggest that differences in attention allocation observed in visual areas may be driven by reward value biases from parietal cortex.

1.3.4 Anatomical connectivity of brain regions involved in value-based attention

In the previous sections, we have seen evidence that suggests that a diverse set of brain regions is involved in determining and implementing the allocation of attention under different conditions. Indeed, most of these brain regions are heavily interconnected. I will briefly outline the connectivity of those brain regions that were explicitly mentioned in the previous introduction.

Dopamine neurons in the midbrain are predominantly located in the ventral tegmental area (VTA) and substantia nigra pars reticulata (SNr) (Francois et al., 1985; Haber et al., 1995b). These regions project densely to subregions of the striatum (ventral striatum, caudate nucleus, putamen) and receive reciprocal complex connections back from the striatum as part of the striato-nigro-striatal network (Haber and Knutson, 2010). Aside from the striatum, dopamine neurons also have diffuse projections to frontal, parietal, temporal and visual cortex (Lidow et al., 1991; Gaspar et al., 1992; Haber and Knutson, 2010), as well as to the amygdala and hippocampus. The main input to midbrain dopamine neurons comes from the striatum and other basal ganglia regions, in addition to sparser projections from the amygdala. The striatum itself, aside from being heavily interconnected with the midbrain dopamine neurons, receives direct and topographical projections from the prefrontal cortex, whereby vmPFC and OFC predominantly project to the ventral striatum, and dorsal ACC and PFC project predominantly to dorsal regions of the striatum (caudate and putamen) (Haber et al., 1995a, 2006; Haber and Knutson, 2010). The striatum projects via the ventral pallidum and thalamus back to the cortex in the fronto-striatal loops. The heavy connectivity between midbrain, basal ganglia and frontal regions may underlie the strong representation of reward and prediction error signals found in those areas. Anatomical regions within the frontal cortex, including dorsal PFC, vmPFC, ACC and OFC are heavily interconnected with each other (Yeterian et al., 2012), and the nature of these connections can be excitatory and inhibitory (Medalla and Barbas, 2009).

These frontal regions are additionally reciprocally connected with a vast set of regions in the temporal cortex, parietal cortex and higher visual cortex (Yeterian et al., 2012). Frontal and parietal cortical areas, including those previously strongly implicated in being part of the attention network (Corbetta and Shulman, 2002), are strongly interconnected (Cavada and Goldman-Rakic, 1989) and in turn have vast subcortical connections to various thalamic nuclei, the amygdala, superior colliculus, as well as connections to insular and rhinal cortex (Selemon and Goldman-Rakic, 1988; Bunce and Barbas, 2011; Bunce et al., 2013). Parietal cortex, including area LIP, is additionally heavily interconnected with areas of the visual extrastriate cortex (Blatt et al., 1990; Lewis and Van Essen, 2000). This suggests that the decision and implementation of where and what we attend to in our environment is realized by a wide network of interconnected brain regions.

1.4 Scope of this dissertation

This dissertation investigates the neural mechanisms that support successful shifts of attention under conditions of correct action selection and under conditions of learning, in macaque monkeys and in humans, (1) by probing the interactions of single neurons in macaque frontal cortex when attention needs to be shifted to a cued target stimulus, (2) by investigating the encoding of feature-specific prediction errors in macaque fronto-striatal areas when the attentional target needs to be actively learned, and (3) by investigating attentional deployment over human visual cortex, and outcome encoding over human frontal cortex, when the attentional target needs to be actively learned.

With this dissertation, I seek to fill several gaps in the existing literature with regards to the neural functions that underlie the decision and implementation of what and where we direct our attention to. In **Chapter 2**, I investigated how single neurons in the macaque frontal cortex functionally interact during covert shifts of attention to a cued target stimulus. As reviewed above, subregions of the frontal cortex, including ACC, are heavily interconnected (Section 1.3.4) and are likely involved in mediating aspects of

cued attention shifts to relevant stimuli (Sections 1.1.1, 1.3.1), and we therefore hypothesized that interactions between some subsets of these regions could be observable between single neurons at the time of attention shifts. Indeed, I found that neuron pairs located in ACC and lateral PFC, specifically, correlate their spike-trains during active shifts of attention and thereby potentially support the successful implementation of covert shifts of attention to one of two target stimuli. This finding complements existing literature that shows functional interactions at the level of scalp potentials or local field potentials in frontal cortical areas during cognitive and attentional control (Tsujimoto et al., 2006; Cavanagh et al., 2009; Rothé et al., 2011; Womelsdorf et al., 2014) and extend these to the single neuron level.

In **Chapters 3** and **4**, I created uncertain environments, using feature-based reversal learning tasks, in which the rewarded feature repeatedly changes, requiring repeated re-learning of attentional targets. This allowed probing of those neural networks that decide what is most valuable to attend to in the environment. As reviewed previously, frontal, parietal and striatal brain regions seem to be involved in learning where to attend - indeed, multiple of those brain regions have been shown to encode various correlates of prediction errors thought to govern learning (Hare et al., 2008; Asaad and Eskandar, 2011; Daw et al., 2011; Hayden et al., 2011; Kennerley et al., 2011, Section 1.3.2). We hypothesized that some of these brain regions may selectively encode prediction error signals informative of preceding feature choices to potentially support the correct assignment of the source of an outcome during active learning. As such, in **Chapter 3**, I used a reinforcement learning approach to identify and characterize neural prediction error signals in frontal, cingulate, and striatal areas of the macaque that carried information about the specific feature that was attended in the preceding choice. I found that feature-specific prediction error signals are ubiquitously represented across those areas that have previously been shown to encode prediction errors, and that those prediction errors that are informative about the valence of the outcome are also most often informative about the relevant feature dimension. These signals may constitute one way the brain solves aspects of the credit assignment problem.

In humans experiencing a similarly uncertain environment, we hypothesized that successful feature-based reversal learning would lead to changes in neural activity correlated with the encoding of prediction errors and attentional target selection. Therefore, in **Chapter 4**, I investigated how changes in attentional and feedback processing, as a function of feature value learning, are reflected in two human event-related potential components: the N2pc, which has been shown to index attentional target selection (Woodman and Luck, 2003; Eimer, 2014), and the FRN, which has been suggested to reflect prediction error encoding (Walsh and Anderson, 2011). I found that N2pc and FRN amplitudes were selectively increased during learning of current value contingencies, and N2pc amplitudes were greater following negative feedback compared with positive feedback. These findings extend the current human EEG and behavioral literature concerned with value-based changes in attention (Della Libera and Chelazzi, 2006; Anderson, 2013; Hickey et al., 2015; Feldmann-Wüstefeld et al., 2016), by highlighting how repeated learning-related changes in attention and feedback processing could support flexible adjustments in behavior.

1.5 References

Alexander WH, Brown JW (2011) Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14:1338–1344.

Anderson BA (2013) A value-driven mechanism of attentional selection. J Vis 13:1–16.

Anderson BA, Laurent PA, Yantis S (2011) Value-driven attentional capture. Proc Natl Acad Sci USA 108:10367–10371.

Apicella P, Deffains M, Ravel S, Legallet E (2009) Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. Eur J Neurosci 30:515–526.

Asaad WF, Eskandar EN (2011) Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J Neurosci 31:17772–17787.

Awh E, Belopolsky A V., Theeuwes J (2012) Top-down versus bottom-up attentional control: A failed theoretical dichotomy. Trends Cogn Sci 16:437–443.

Ball K, Sekuler R (1980) Models of stimulus uncertainty in motion perception. Psychol Rev 87:435:469.

Balleine BW, Liljeholm M, Ostlund SB (2009) The integrative function of the basal ganglia in instrumental conditioning. Behav Brain Res 199:43–52.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Bichot NP, Schall JD (1999) Effects of similarity and history on neural mechanisms of visual selection. Nat Neurosci 2:549–554.

Bichot NP, Schall JD (2002) Priming in macaque frontal cortex during popout visual search: feature-based facilitation and location-based inhibition of return. J Neurosci 22:4675–4685.

Blanchard TC, Hayden BY, Bromberg- ES, Correspondence M, Bromberg-Martin ES (2015) Orbitofrontal Cortex Uses Distinct Codes for Different Choice Attributes in Decisions Motivated by Curiosity. Neuron 85:602–614.

Blatt GJ, Andersen RA, Stoner GR (1990) Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. J Comp Neurol 299:421–445.

Bromberg-Martin ES, Hikosaka O (2009) Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron 63:119–126.

Brosch T, Sander D, Pourtois G, Scherer KR (2008) Beyond Fear - Rapid Spatial Orienting Toward Positive Emotional Stimuli. Psychol Sci 19:362–370.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PGF, Kwok SC, Phillips A, Tanaka K (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science 325:52–58.

Buffalo EA, Fries P, Landman R, Liang H, Desimone R (2010) A backward progression of attentional effects in the ventral stream. Proc Natl Acad Sci U S A 107:361–365.

Bunce JG, Barbas H (2011) Prefrontal pathways target excitatory and inhibitory systems in memory-related medial temporal cortices. Neuroimage 55:1461–1474.

Bunce JG, Zikopoulos B, Feinberg M, Barbas H (2013) Parallel prefrontal pathways reach distinct excitatory and inhibitory systems in memory-related rhinal cortices. J Comp Neurol 521:211–220.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860–1862.

Cavada C, Goldman-Rakic PS (1989) Posterior parietal cortex in rhesus monkeys: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. J Comp Neurol 28:422–445.

Cavanagh JF, Cohen MX, Allen JJB (2009) Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. J Neurosci 29:98–105.

Chau BKH, Sallet J, Papageorgiou GK, Noonan MAP, Bell AH, Walton ME, Rushworth MFS (2015) Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. Neuron 87:1106–1118.

Chun MM (2000) Contextual cueing of visual attention. Trends Cogn Sci 4:170–178.

Chun MM, Jiang Y (1998) Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. Cogn Psychol 36:28–71.

Chun MM, Phelps E a (1999) Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nat Neurosci 2:844–847.

Chun MM, Turk-Browne NB (2007) Interactions between attention and memory. Curr Opin Neurobiol 17:177–184.

Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594–1600.

Colby CL, Duhamel JR, Goldberg ME (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. J Neurophysiol 76:2841–2852.

Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. J Neurophysiol 75:1306–1308.

Corbetta M, Shulman GL (2002) Control of Goal-Directed and Stimulus-Driven Attention in the Brain. Nat Rev Neurosci 3:215–229.

Corlett PR, Aitken MRF, Dickinson A, Shanks DR, Honey GD, Honey RAE, Robbins TW, Bullmore ET, Fletcher PC (2004) Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. Neuron 44:877–888.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.

Dayan P, Balleine BW (2002) Reward, Motivation, and Reinforcement Learning. Neuron 36:285–298.

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cogn Affect Behav Neurosci 8:429–453.

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3:1218–1223.

Dayan P, Niv Y (2008) Reinforcement learning: The Good, The Bad and The Ugly. Curr Opin Neurobiol 18:185–196.

Della Libera C, Chelazzi L (2006) Visual selective attention and the effects of monetary rewards. Psychol Sci 17:222–227.

Della Libera C, Chelazzi L (2009) Learning to attend and to ignore is a matter of gains and losses. Psychol Sci 20:778–784.

den Ouden HEM, Kok P, de Lange FP (2012) How prediction errors shape perception, attention, and

motivation. Front Psychol 3:1–12.

Den Ouden HEM, Kok P, de Lange FP (2012) How prediction errors shape perception, attention, and motivation. Front Psychol 3:1–12.

Desimone R, Duncan J (1995) Neural Mechanisms of Selective Visual. Annu Rev Neurosci 18:193–222.

Dias R, Robbins TW, Roberts AC (1996) Dissociation in prefrontal cortex of affective and attentional shifts. Nature 380:69–72.

Dickinson A (1980) Contemporary Animal Learning Theory. Cambridge, MA: Cambridge University Press.

Dopson JC, Esber GR, Pearce JM (2010) Differences in the associability of relevant and irrelevant stimuli. J Exp Psychol Anim Behav Process 36:258–267.

Dorris MC, Klein RM, Everling S, Munoz DP (2002) Contribution of the primate superior colliculus to inhibition of return. J Cogn Neurosci 14:1256–1263.

Eimer M (2014) The neural basis of attentional control in visual search. Trends Cogn Sci 18:526–535.

Eimer M, Kiss M, Cheung T (2010) Priming of pop-out modulates attentional target selection in visual search: Behavioural and electrophysiological evidence. Vision Res 50:1353–1361.

Engelmann JB, Pessoa L (2007) Motivation sharpens exogenous spatial attention. Emotion 7:668–674.

Eriksen CW, Hoffman JE (1973) The extent of processing of noise elements during selective encoding from visual displays. Percept Psychophys 14:155–160.

Fecteau JH, Bell AH, Munoz DP (2004) Neural Correlates of the Automatic and Goal-Driven Biases in Orienting Spatial Attention. J Neurophysiol 92:1728–1737.

Feldmann-Wüstefeld T, Brandhofer R, Schubö A (2016) Rewarded visual items capture attention only in heterogeneous contexts. Psychophysiology 53:1063–1073.

Ferrera VP, Rudolph KK, Maunsell JH (1994) Responses of neurons in the parietal and temporal visual pathways during a motion task. J Neurosci 14:6171–6186.

Field M, Mogg K, Bradley BP (2004) Eye movements to smoking-related cues: Effects of nicotine deprivation. Psychopharmacology (Berl) 173:116–123.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science (80-) 299:1898–1902.

Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, Papadakis N, Bullmore ET (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nat Neurosci 4:1043–1048.

Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J (2017) Parietal neurons encode expected gains in instrumental information. Proc Natl Acad Sci 114.

Francois C, Percheron G, Yelnik J, Heyner S (1985) A histological atlas of the macaque (Macaca, mulatta) substantia nigra in ventricular coordinates. Brain Res Bull 14:349–367.

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. Science 291:1560–1563.

Friston KJ (2005) A theory of cortical responses. Philos Trans R Soc B 360:815–836.

Gaspar P, Stepniewska I, Kaas JH (1992) Topography and collateralization of the dopaminergic projections to motor and lateral prefrontal cortex in owl monkeys. J Comp Neurol 325:1–21.

George DN, Pearce JM (1999) Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. J Exp Psychol Anim Behav Process 25:363–373.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.

Gottlieb J (2012) Attention, learning and the value of information. Neuron 76:281–295.

Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. Science (80-) 324:1207–1210.

Guo R, Bohmer W, Hebart M, Chien S, Sommer T, Obermayer K, Gläscher J (2016) Interaction of Instrumental and Goal-Directed Learning Modulates Prediction Error Representations in the Ventral Striatum. J Neurosci 36:12650–12660.

Haber SN, Kim K-S, Mailly P, Calzavara R (2006) Reward-Related Cortical Inputs Define a Large Striatal Region in Primates That Interface with Associative Cortical Connections, Providing a Substrate for Incentive-Based Learning. J Neurosci 26:8368–8376.

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26.

Haber SN, Kunishio K, Mizobuchi M (1995a) The Orbital and Medial Prefrontal Basal Ganglia Circuit Through the Primate. J Neurosci 15:4851–4867.

Haber SN, Ryoo H, Cox C, Lu W (1995b) Subsets of Midbrain Dopaminergic-Neurons in Monkeys are Distinguished by Different Levels of Messenger-Rna for the Dopamine Transporter - Comparison with the Messenger-Rna for the D-2 Receptor, Tyrosine-Hydroxylase and Calbindin Immunoreactivity. J Comp Neurol 362:400–410.

Haenny PE, Schiller PH (1988) State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Exp brain Res 69:225–244.

Hall G, Pearce JM (1979) Latent inhibition of a CS during CS-US pairings. J Exp Psychol Anim Behav Process 5:31–42.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci 28:5623–5630.

Haselgrove M, Esber GR, Pearce JM, Jones PM (2010) Two kinds of attention in Pavlovian conditioning: evidence for a hybrid model of learning. J Exp Psychol Anim Behav Process 36:456–470.

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. J Neurosci 31:4178–4187.

Hickey C, Chelazzi L, Theeuwes J (2010) Reward changes salience in human vision via the anterior cingulate. J Neurosci 30:11096–11103.

Hickey C, Kaiser D, Peelen M V (2015) Reward Guides Attention to Object Categories in Real-World Scenes. J Exp Psychol Gen 144:264–273.

Hogarth L, Dickinson A, Austin A, Brown C, Duka T (2008) Attention and expectation in human predictive learning: the role of uncertainty. Q J Exp Psychol (Hove) 61:1658–1668.

Hogarth L, Dickinson A, Duka T (2010) Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal, and uncertainty. In: Attention and Associative Learning. From Brain to Behavior (Mitchell CJ, Le Pelley

ME, eds), pp 71–97. Oxford University Press.

Holland PC, Gallagher M (1999) Amygdala circuitry in attentional and representational processes. Trends Cogn Sci 3:65–73.

Kaping D, Vinck M, Hutchison RM, Everling S, Womelsdorf T (2011) Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. PLoS Biol 9:e1001224.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315–341.

Kennerley SW, Behrens TEJ, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581–1589.

Kok P, Rahnev D, Jehee JFM, Lau HC, De Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206.

Kristjansson A, Campana G (2010) Where perception meets memory: A review of repetition priming in visual search tasks. Atten Percept Psychophys 72:5–18.

Kristjansson A, Vuilleumier P, Schwartz S, Macaluso E, Driver J (2007) Neural Basis for Priming of Pop-Out during Visual Search Revealed with fMRI. Cereb Cortex 17:1612–1624.

Le Pelley ME (2004) The role of associative history in models of associative learning: a selective review and a hybrid model. Q J Exp Psychol B 57:193–243.

Lee D, Seo H, Jung MW (2012) Neural Basis of Reinforcement Learning and Decision Making. Annu Rev Neurosci 35:287–308.

Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y (2017) Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron 93:451–463.

Lewis JW, Van Essen DC (2000) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J Comp Neurol 428:112–137.

Li L, Gratton C, Yao D, Knight RT (2010) Role of frontal and parietal cortices in the control of bottom-up and top-down attention in humans. Brain Res 1344:173–184.

Lidow MS, Goldman-Rakic PS, Gallager DW, Rakic P (1991) Distribution of dopaminergic receptors in the primate cerebral cortex: Quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390. Neuroscience 40:657–671.

Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci U S A 104:9493–9498.

Luck SJ, Chelazzi L, Hillyard S a, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex. J Neurophys 77:24–42.

Lupiáñez J, Klein RM, Bartolomeo P (2006) Inhibition of return: Twenty years after. Cogn Neuropsychol 23:1003–1014.

Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev 82:276–298.

Mackintosh NJ (1976) Overshadowing and stimulus intensity. Anim Learn Behav 4:186–192.

Mackintosh NJ (1983) Conditioning and Associative Learning. Oxford: Oxford University Press.

Mackintosh NJ, Little L (1969) Intradimensional and extradimensional shift learning by pigeons. Psychon Sci 14:5–6.

Maddux J-M, Kerfoot E, Chatterjee S, Holland PC (2007) Dissociation of attention in learning and action: Effects of lesions of the amygdala central nucleus, medial prefrontal cortex, and posterior parietal cortex. Behav Neurosci 121:63–79.

Maddux JM, Holland P (2011) Dissociations between medial prefrontal cortical subregions in the modulation of learning and action. Behav Neurosci 125:383–395.

Maljkovic V, Nakayama KEN (1994) Priming of pop-out : I . Role of features. 22:657–672.

Maljkovic V, Nakayama KEN (1996) Priming of pop-out : II . The role of position. 58:977–991.

Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447:1111–1115.

Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647–656.

Maunsell JHR (2004) Neuronal representations of cognitive state: Reward or attention? Trends Cogn Sci 8:261–265.

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441.

Medalla M, Barbas H (2009) Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. Neuron 61:609–620.

Miller EK, Buschman TJ (2012) Cortical circuits for the control of attention. Curr Opin Neurobiol:1–7.

Mitchell JF, Sundberg K a, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55:131–141.

Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63:879–888.

Mogg K, Bradley BP, Hyare H, Lee S (1998) Selective attention to food-related stimuli in hunger: Are attentional biases specific to emotional and psychopathological states, or are they also found in normal drive states? Behav Res Ther 36:227–237.

Morris RW, Bouton ME (2006) Effect of unconditioned stimulus magnitude on the emergence of conditioned responding. J Exp Psychol Anim Behav Process 32:371–385.

Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol 70:909–919.

Motter BC (1994) Neural Correlates of Attentive Selection for Color or Luminance in Extrastriate Area V4. J Neurosci 14:2178–2189.

Navalpakkam V, Koch C, Rangel A, Perona P (2010) Optimal reward harvesting in complex perceptual environments. Proc Natl Acad Sci U S A 107:5232–5237.

Niv Y (2009) Reinforcement learning in the brain. J Math Psychol 53:139–154.

Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J Neurosci 30:10692–10702.

O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci U S A 110:E3660-9.

Oemisch M, Westendorff S, Everling S, Womelsdorf T (2015) Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. J Neurosci 35:13076–

13089.

Pavlov IP (1927) “Conditioned reflexes”. An investigation of the physiological activity of the cerebral cortex. Oxford: Oxford University Press.

Pearce J, Hall G (1980) A model for Pavlovian learning: Variation in the effectiveness of conditioned but not unconditioned stimuli. Psychol Rev 87:532–552.

Pearce J, Mackintosh NJ (2010) Two theories of attention: a review and possible integration. In: Attention and Associative Learning. From Brain to Behavior (Mitchell CJ, Le Pelley ME, eds), pp 11–39. Oxford University Press.

Pearce JM, Hall G (1979) Loss of associability by a compound stimulus comprising excitatory and inhibitory elements. J Exp Psychol Anim Behav Process 5:19–30.

Peck CJ, Jangraw DC, Suzuki M, Efem R, Gottlieb J (2009) Reward Modulates Attention Independently of Action Value in Posterior Parietal Cortex. J Neurosci 29:11182–11191.

Pessoa L, Engelmann JB (2010) Embedding reward signals into perception and cognition. Front Neurosci 4:1–8.

Pflugshaupt T, Mosimann UP, Von Wartburg R, Schmitt W, Nyffeler T, Müri RM (2005) Hypervigilance-avoidance pattern in spider phobia. J Anxiety Disord 19:105–116.

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25.

Posner MI, Cohen Y (1984) Components of visual orienting. Atten Perform Control Lang Process 32:531–556.

Posner MI, Rafal RD, Choate LS, Vaughan J (1985) Inhibition of Return: Neural Basis and Function. Cogn Neuropsychol 2:211–228.

Raymond M. Klein (2000) Inhibition of return. Trends Cogn Neurosci 4:138–147.

Rescorla R, Wagner A (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.J., Prokasy,W.F. (Eds.), Classical Conditioning. II. Current Research and Theory. Appelton-Century Crofts, New York, pp 64–99.

Reynolds JH, Heeger DJ (2009) The Normalization Model of Attention. Neuron 61:168–185.

Rossi AF, Pessoa L, Desimone R, Ungerleider LG (2009) The prefrontal cortex and the executive control of attention. Exp Brain Res 192:489–497.

Rothé M, Quilodran R, Sallet J, Procyk E (2011) Coordination of high gamma activity in anterior cingulate and lateral prefrontal cortical areas during adaptation. J Neurosci 31:11110–11117.

Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal Cortex and Reward-Guided Learning and Decision-Making. Neuron 70:1054–1069.

Saalmann YB, Pigarev IN, Vidyasagar TR (2007) Neural Mechanisms of Visual Attention: How Top-Down Feedback Highlights Relevant Locations. Science (80-) 316:1612–1615.

Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. Science 337:753–756.

San Martín R, Appelbaum LG, Huettel SA, Woldorff MG (2016) Cortical Brain Activity Reflecting Attentional Biasing Toward Reward-Predicting Cues Covaries with Economic Decision-Making Performance. Cereb Cortex 26:1–11.

Sapir A, Soroker N, Berger A, A. H (1999) Inhibition of return in spatial attention: direct evidence for collicular generation. Nat Neurosci 2:1053–1054.

Schacter DL, Buckner RL (1998) Priming and the Brain. Neuron 20:185–195.

Schultz W, Apicella P, Ljungberg T (1993) Responses of Monkey Dopamine Neurons to Reward and Conditioned Stimuli during Successive Steps of Learning a Delayed Response Task. J Neurosci 13:900–913.

Seitz AR, Kim D, Watanabe T (2009) Rewards Evoke Learning of Unconsciously Processed Visual Stimuli in Adult Humans. Neuron 61:700–707.

Selemon LD, Goldman-Rakic PS (1988) Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. J Neurosci 8:4049–4068.

Serences JT, Yantis S (2006) Selective visual attention and perceptual coherence. Trends Cogn Sci 10:38–45.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Shuler MG, Bear MF (2006) Reward Timing in the Primary Visual Cortex. Science (80-) 311:1606–1609.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2012) Reward prediction error signaling in posterior dorsomedial striatum is action-specific. J Neurosci 32:10296–10305.

Sul JH, Kim H, Huh N, Lee D, Jung MW (2010) Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron 66:449–460.

Summerfield C, Koechlin E (2008) A Neural Representation of Prior Information during Perceptual Inference. Neuron 59:336–347.

Summerfield JJ, Lepsien J, Gitelman DR, Mesulam MM, Nobre AC (2006) Orienting attention based on long-term memory experience. Neuron 49:905–916.

Sutton RS, Barto A (1998) Reinforcement Learning. Cambridge, MA: MIT Press.

Tatler BW, Hayhoe MM, Land MF, Ballard DH (2011) Eye guidance in natural vision: Reinterpreting salience. J Vis 11:5–5.

Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science (80-) 307:1642–1645.

Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399:575–579.

Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. Nature 382:539–541.

Tsujimoto T, Shimazu H, Isomura Y (2006) Direct recording of theta oscillations in primate prefrontal and anterior cingulate cortices. J Neurophysiol 95:2987–3000.

Vickers JN (2007) Perception, Cognition, and decision training: The quiet eye in action. Human Kinetics.

Vickers JN (2016) Origins and current issues in Quiet Eye research. Curr Issues Sport Sci 1:1–11.

Voloh B, Valiante TA, Everling S, Womelsdorf T (2015) Theta-gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. Proc Natl Acad Sci U S A 112:8457–8462.

Vuilleumier P (2005) How brains beware: Neural mechanisms of emotional attention. Trends Cogn Sci 9:585–594.

Vuilleumier P, Armony JL, Driver J, Dolan RJ (2003) Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat Neurosci 6:624–631.

Wacongne C, Labyt E, Wassenhove V van, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci 108:20754–20759.

Walsh MM, Anderson JR (2011) Modulation of the feedback-related negativity by instruction and experience. Proc Natl Acad Sci USA 208:19048–19035.

Walton ME, Behrens TEJ, Noonan MP, Rushworth MFS (2011) Giving credit where credit is due : orbitofrontal cortex and valuation in an uncertain world. Ann N Y Acad Sci 1239:14–24.

Weierich MR, Treat TA, Hollingworth A (2008) Theories and measurement of visual attentional processing in anxiety.

Wilson PN, Boumphrey P, Pearce JM (1992) Restoration of the orienting response to a light by a change in its predictive accuracy. J Exp Psychol 44:17–36.

Womelsdorf T, Ardid S, Everling S, Valiante TA (2014) Burst Firing Synchronizes Prefrontal and Anterior Cingulate Cortex during Attentional Control. Curr Biol:1–9.

Womelsdorf T, Everling S (2015) Long-Range Attention Networks: Circuit Motifs Underlying Endogenously Controlled Stimulus Selection. Trends Neurosci 38:682–700.

Womelsdorf T, Fries P, Mitra PP, Desimone R (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. Nature 439:733–736.

Woodman GF, Luck SJ (2003) Serial deployment of attention during visual search. J Exp Psychol Hum Percept Perform 29:121–138.

Yacubian J (2006) Dissociable Systems for Gain- and Loss-Related Value Predictions and Errors of Prediction in the Human Brain. J Neurosci 26:9530–9537.

Yeterian EJ, Pandya DN, Tomaiuolo F, Petrides M (2012) The cortical connectivity of the prefrontal cortex in the monkey brain. Cortex 48:58–81.

Chapter 2

Inter-areal Spiketrain Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts

Mariann Oemisch[1], Stephanie Westendorff[1], Stefan Everling[2], Thilo Womelsdorf[1]

[1]Department of Biology, Centre for Vision Research, York University, Toronto, Ontario M6J 1P3, Canada; [2]Department of Physiology and Pharmacology, Centre for Functional and Metabolic Mapping, Western University, Ontario N6A 5B7, Canada

Journal of Neuroscience, 2015, 35(38): 13076-13089

2.1 Summary

Anterior cingulate and prefrontal cortex (ACC/PFC) are believed to co-activate during goal-directed behavior to identify, select and monitor relevant sensory information. Here we tested whether co-activation of neurons across macaque ACC/PFC would be evident at the level of pairwise neuronal correlations during stimulus selection in a spatial attention task. We found that firing correlations emerged shortly after an attention cue, were evident for 50-200ms time windows, strongest for neuron pairs in area 24 (ACC) and areas 8 and 9 (dorsal PFC), and independent of overall firing rate modulations. For a subset of cell pairs from ACC and dorsal PFC the observed functional spiketrain connectivity carried information about the direction of the attention shift. Reliable firing correlations were evident across area boundaries for neurons with broad spike waveforms (putative excitatory neurons) as well as for pairs of putative excitatory neurons and neurons with narrow spike waveform (putative interneurons). These findings reveal that stimulus selection is accompanied by slow time scale firing correlations across those ACC/PFC subfields implicated to control and monitor attention. This functional coupling was informative about which stimulus was selected and thus indexed possibly the exchange of task-relevant information. We speculate that interareal, transient firing correlations reflect the transient coordination of larger, reciprocally interacting brain networks at a characteristic 50-200ms time scale.

2.2 Introduction

Multiple anatomically segregated brain areas play a role in the flexible allocation of attention during goal-directed behavior. Among those areas are the anterior cingulate cortex (ACC) and the prefrontal cortex (Buckley et al., 2009; Miller and Buschman, 2012; Passingham and Wise, 2012; Genovesio et al., 2014). The ACC is suggested to monitor and guide attentional selection, while lateral and dorsal prefrontal areas are involved in the actual implementation of attentional control, and ventromedial prefrontal areas

(vmPFC) contribute relevant value estimates of possible attentional targets (Alexander and Brown, 2011; Kaping et al., 2011; Kennerley et al., 2011; Shenhav et al., 2013). To this point it is unclear how attention information is integrated across these distributed brain areas, and studies investigating direct, inter-areal correlations between these areas have been sparse. One possible mechanism is that neurons across these brain areas synchronize their firing to delegate information. The communication between the PFC subareas and the ACC has been tested by few studies that employed electroencephalography (EEG) (Cavanagh et al., 2009) or local field potential (LFP) measures (Tsujimoto et al., 2006; Rothé et al., 2011; Womelsdorf et al., 2014a). For instance, increased theta band phase coherence between ACC and lateral PFC electrodes has previously been found in humans and macaques performing different tasks that probed cognitive control and task-switching abilities (Cavanagh et al., 2009; Phillips et al., 2013). Recent findings in the macaque suggest that the interaction between lateral PFC areas and ACC can be traced back to the burst firing events of cells in ACC, with bursts in the ACC synchronizing to the LFP in PFC during attentional states (Womelsdorf et al., 2014a). These findings provide compelling evidence for the existence of direct communication between ACC and PFC during attentional control. This interaction may also be evident in functional firing correlations between single cells in both brain areas given their largely dendrite mediated interconnectivity (e.g. Medalla and Barbas, 2009).

Multiple methods exist for analyzing spiketrain correlations between single cells, one being the joint peristimulus time histogram (JPSTH) analysis (Aertsen et al., 1989; Vaadia et al., 1995b; Paz et al., 2007; Joshua et al., 2009; Schultz et al., 2009; Cohen et al., 2010, Zhuang et al., 2014), which we utilized in this work. The JPSTH allows a time-resolved analysis of the dynamic correlation of the spikes of one neuron with those of a second neuron while controlling for an influence of overall firing rate modulations. Here, we used JPSTH analysis to test whether neuron pairs across areas of the ACC/PFC transiently coordinate by correlating their firing when attention shifted to one of two possible target stimuli. We found that cells on average increased their spiketrain correlations approximately 280 ms after an attention cue that triggered the covert selection of one peripheral stimulus. Spiketrain correlations occurred on a slow 50-

200ms time scale and between distant neurons and are therefore unlikely to be caused by direct monosynaptic connections. These correlations observed are therefore qualitatively different from short (<5ms) time-lag cross correlations of neurons recorded within close (<~0.4mm) proximity (e.g. Kohn and Smith, 2005; Zhou et al., 2014). The emergence of attention specific spiketrain correlations was not dependent on overall firing rate modulations, emerged with anatomical specificity, and was found between neurons in ACC and dorsal PFC, where spiketrain correlations carried attention information about the location of the relevant target stimulus. Furthermore, spiketrain correlations could be identified in pairs of isolated single neurons with broad as well as narrow waveforms (putative pyramidal neurons and interneurons, respectively), providing a direct window onto how individual neurons contribute to larger recurrent network activations at slow, 50-200ms time scales.

2.3 Methods

2.3.1 Procedures

Data was collected from two male rhesus macaques (*Macaca mulatta*). All animal care and experimental protocols were approved by the University of Western Ontario Council on Animal Care and were in accordance with the Canadian Council on Animal Care guidelines. Details regarding the experimental set-up, recording procedures and reconstruction of recording sites have been described previously (Kaping et al., 2011). Briefly, extra-cellular recordings were made with 1 - 6 tungsten electrodes (impedance 1.2 - 2.2 MOhm, FHC, Bowdoinham, ME) through standard recording chambers (19 mm inner diameter) implanted over the left hemispheres in both monkeys. Electrodes were lowered daily through guide tubes and a recording grid with 1 mm inter-hole spacing using software controlled precision micro-drives (NAN Instruments Ltd., Israel). Data amplification, filtering, and acquisition were done with a multi-channel acquisition processor (Map System, Plexon, Inc.). Spiking activity was obtained following a 100 - 8,000 Hz passband filter and further amplification and digitization at 40 kHz sampling rate. Eye positions were

monitored using a video-based eye-tracking system (ISCAN, Woburn, US, sampling rate: 120 Hz) calibrated prior to each experiment to a 5-point fixation pattern (one central fixation point and the remaining four points offset by vertical 8.8º and horizontal 6º toward the 4 corners of the monitor). Eye fixation was controlled within a 1.4-2.0 degree radius window. Sorting and isolation of single unit activity was performed offline with Plexon Offline Sorter (Plexon Inc., Dallas, TX), based on principal component analysis of the spike waveforms.

During the experiments, stimulus presentation, monitored eye positions, and reward delivery were controlled via MonkeyLogic (open-source software http://www.brown.edu/Research/monkeylogic/).

2.3.2 Behavioral paradigm

Monkeys performed a selective attention task requiring a two-alternative forced-choice discrimination (Figure 2.1A). The behavioral task has previously been described in detail (Kaping et al., 2011). Briefly, each trial started with the appearance of a grey central fixation point, which monkeys had to fixate. After 0.3s, two black/white drifting gratings appeared to the left and right of the central fixation point. Following another 0.4s the two stimuli gratings changed color to black/green and black/red. After 0.05 - 0.75s the central fixation point turned either green or red instructing the monkey to covertly shift attention to the matching stimulus grating. Another 0.05 - 4s later, both stimulus gratings rotated with a temporal gap either clockwise or counter-clockwise. The monkey had to discriminate the clockwise or counter-clockwise rotation of the cued stimulus grating and respond with an up- or downward saccade respectively to one of two target dots. In fifty percent of the trials, the non-cued stimulus grating rotated first which had to be ignored by the monkey. A correct saccadic response had to occur within 0.05 and 0.6 s following rotation onset to a target dot vertically above or below the fixation point. The response target dot had to be fixated for a minimum of 50ms, which then resulted in a liquid reward. Reward size varied between 0.76 and 0.4 ml for different colors across blocks of 30 trials. Reward associations changed

across blocks and were equally balanced and are therefore expected to not systematically affect any analysis we report and conclusions we draw.

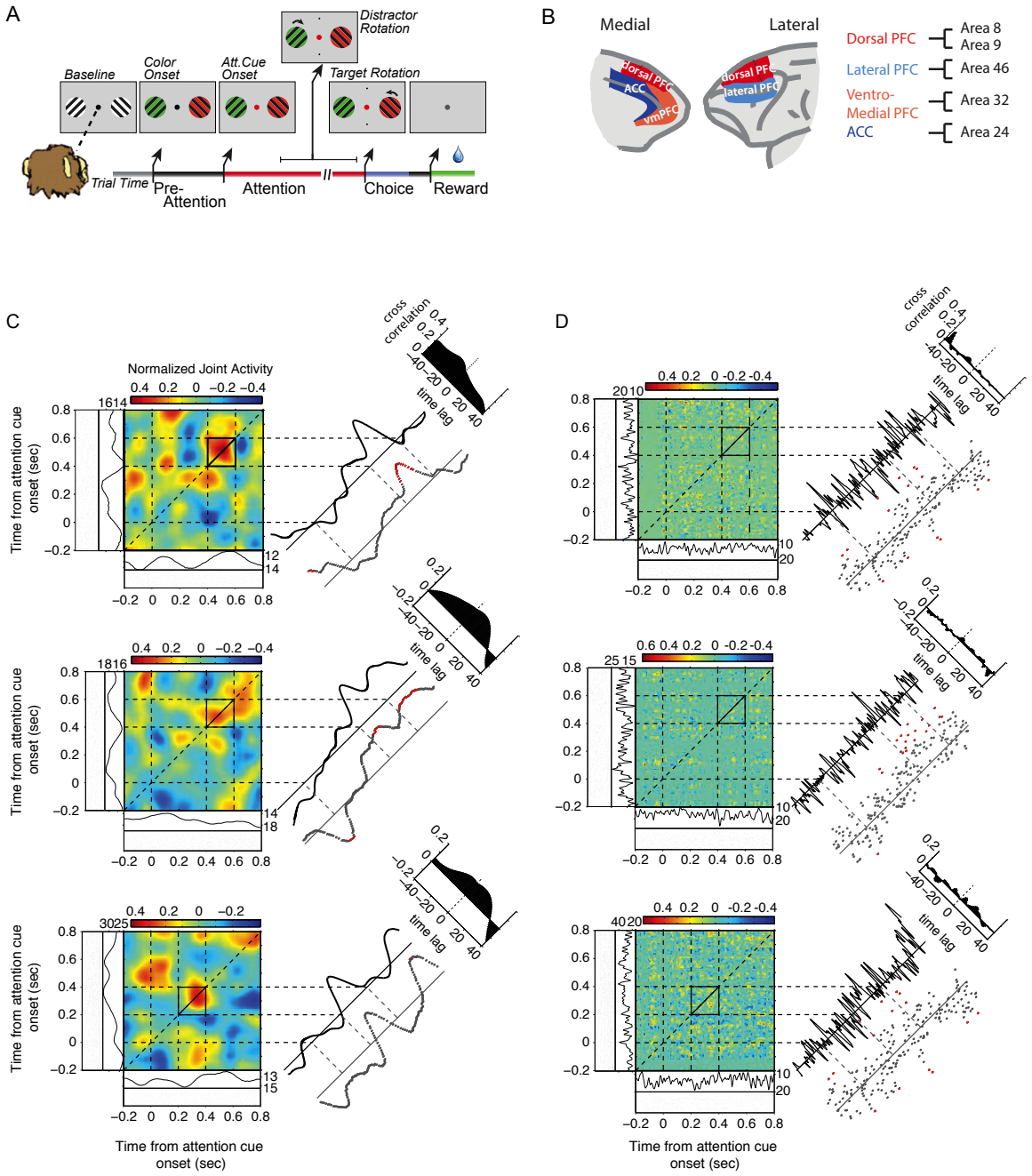

Figure 2.1 Task, Recording Locations and Example JPSTHs.

(A) The selective attention task required monkeys to fixate a central point, while presented with two peripheral grating stimuli. First, both grating stimuli changed their color to either green or red. Then, the

fixation point changed its color to match one of the two gratings, and the monkeys had to use this cue instruction to covertly attend the relevant stimulus. Monkeys then sustained focused attention until the relevant stimulus rotated, filtering out a potential rotation of the distractor stimulus, and finally report it with a saccadic response to one of two target locations (top vs. bottom) in association to a clockwise vs. counterclockwise rotation. Monkeys received reward when correctly indicating the rotation direction of the cued stimulus. See Material and Methods for more details. **(B)** The medial (left) and lateral (right) prefrontal cortex of the macaques. The colors label the four cortical fields from which neuronal activity was recorded. The cortical subdivision follows anatomical parcellation results by Saleem et al. (2014). **(C)** Three examples of normalized and shiftpredictor-corrected JPSTHs (Aertsen et al., 1989) which were computed with spike density functions with a Gaussian kernel width of 50ms. The line to the immediate right of the JPSTH represents the coincidence line of the JPSTH, averaged over ±25ms around the main diagonal. The top right corner depicts the cross correlogram of the sector of the JPSTH defined by the black square. It was computed by collapsing over the time axis. Below the coincidence line are the Pearson correlation coefficients computed in 5ms windows. Red dots indicate significant correlation (t-test ≤ 0.05). Below and to the left of the JPSTH are the PSTH and raster plot of the two neurons from which the JPSTH was computed. **(D)** Same examples and same conventions as in **C**, but computed with a Gaussian kernel width of 5ms.

2.3.3 Anatomical reconstruction

We recorded from neurons in prefrontal and anterior cingulate cortex across subfields that we identified following anatomical reconstruction as outlined before (Kaping et al. 2011). In summary, we projected each electrode trajectory onto the two-dimensional brain slices obtained from 7T anatomical MRI images, using the open-source OsiriX Imaging software (Rosset et al., 2004) and custom-written MATLAB programs (The Mathworks Inc.), utilizing the iodine visualized electrode trajectory within the electrode grid placed within the recording chamber during MR scanning. We drew the coronal outline of the cortical folding of the MR grey scale image to ease the comparison of the individuals' monkey brain slices to standard anatomical atlases, before projecting the electrode tip position into the standardized macaque brain template (F99) available in Caret (http://www.nitrc.org/projects/caret). Note that we initially reproduced the individual monkey brains within the Caret software to validate similarity and derive the scaling factors to match the lower resolution monkey MRs to the higher resolution standard F99 brain. We then manually projected, under visual guidance, the electrode position to the matched location in the standard brain in Caret (van Essen, 2002). In an independent procedure, we visualized major anatomical subdivision schemes of the fronto-cingulate cortex, using the scheme from the Price lab (Saleem et al., 2008, 2014) as reference. Similar results are obtained when using other

prefrontal/cingulate cortex subdivision schemes from Barbas and Zikopoulos (2007), from Petrides and Pandya (1994, 1999), as well as from Sallet and colleagues (2013). Neurons were recorded in the anterior cingulate cortex (area 24), the dorsal prefrontal cortex (area 8, 9), the lateral prefrontal cortex (area 46) and the ventromedial prefrontal cortex (area 32) (see Figure 2.1B).

2.3.4 Data analysis

Analysis was performed with custom MATLAB code (Mathworks, Natick, MA), utilizing functionality from the open-source fieldtrip toolbox (http://www.ru.nl/fcdonders/fieldtrip/). The analysis of spiking activity was limited to neurons with a minimum average firing rate of 1Hz and a minimum of 30 spikes. Spiketrains were either aligned to the onset of the attention cue (cue-aligned) or to the onset of stimulus color (color-aligned), depending on the analysis. For cue-aligned data, trials in which color onset occurred within 200 ms before attention cue onset, and rotation onset occurred within 1000 ms after attention cue onset, were discarded. For color-aligned data, trials in which attention cue onset occurred within 200 ms after color cue onset were discarded. Unless otherwise specified, spike density functions were convolved using a Gaussian kernel with a standard deviation of 50 ms. We define this kernel width as a time window of 100 ms. The following analysis is therefore based on smoothed spiketrains and cannot be directly compared to spike cross-correlation analyses using unsmoothed data (e.g. Kohn and Smith, 2005; Zhou et al., 2014).

Joint Peri-Stimulus Time Histograms

Spike-timing relations between neuron pairs were measured using joint peristimulus time histograms (JPSTHs) (Aertsen et al., 1989). The normalized JPSTH of neurons i and j was defined as follows:

$$J_{Ni,j}(t_1,t_2) = \frac{< S_i^r(t_1) S_j^r(t_2) >}{\sqrt{\frac{1}{N}\Sigma(S_i(t_1) - \overline{S}_i(t_1))^2 \frac{1}{N}\Sigma(S_j(t_2) - \overline{S}_j(t_2))^2}} \quad \text{(eq. 1)}$$

where $S_j^r(t)$ represents the activity of neuron *j* for the *r*th trial at time t, t1 and t2 are time points within the trial, < > represents averaging over r trials and *N* is the number of trials. $\overline{S}_i(t_1)$ represents the average activity of neuron i at time t over all trials. In order to remove stimulus-induced covariations in firing (de la Rocha et al., 2007), we computed a normalized shift-predictor in consecutive trials in a symmetric way:

$$P_{Ni,j}(t_1,t_2) = \frac{< S_i^r(t_1)S_j^{r-1}(t_2) + S_i^{r-1}(t_1)S_j^r(t_2) >}{2\sqrt{\frac{1}{N}\Sigma(S_i(t_1) - \overline{S}_i(t_1))^2 \frac{1}{N}\Sigma(S_j(t_2) - \overline{S}_j(t_2))^2}} \quad \text{(eq. 2)}$$

with the same conventions as above. Picking consecutive trials and computing the shift predictor in a symmetric way ensures that slow changes in the temporal structure do not affect the shift predictor (as opposed to shuffling the order of all trials for one of two channels). The shift predictor does not correct for excitability fluctuations that are independent of stimulus-induced covariations. The normalized shift predictor was subtracted from the normalized JPSTH in order to produce the final shift-predictor-corrected JPSTH (Figure 2.1C, D). Spike density functions used to compute JPSTHs were smoothed with Gaussian kernels with a standard deviation (SD) of 50ms (Figure 2.1C, Figure 2.2A). For visualization purposes, JPSTHs were also computed using spike density functions smoothed with a Gaussian kernel with an SD of 5ms (Figure 2.1D, Figure 2.2E). The JPSTH analysis, compared to a simpler Pearson correlation analysis (see below), allows an estimate of firing-rate corrected spiketrain-to-spiketrain correlations that is not restricted to only coincident firing, but also allows insight into the timing-relations of these interactions. The diagonal of the JPSTH corresponds to coincident firing and measures the average spike timing correlation over time and will in the following be referred to as coincidence line and as an indicator for the absence or presence of what we refer to as *spiketrain correlations*. The coincidence line was computed by averaging the JPSTH's main diagonal ± 25 ms (Figure 2.1C, D).

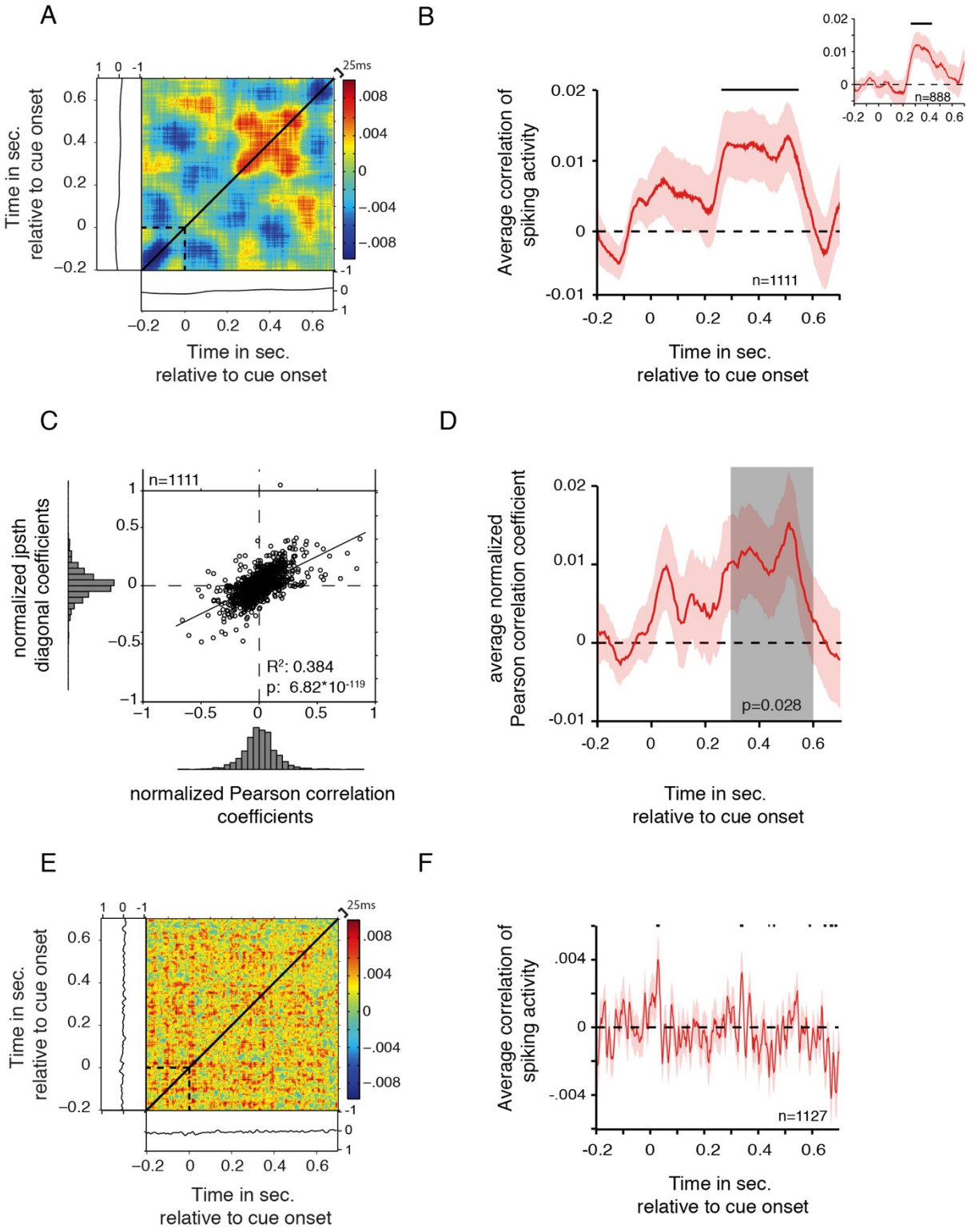

Figure 2.2 Average Firing Correlations with JPSTH and Pearson correlation.

(A) Average normalized JPSTH computed over all task-modulated cell pairs. Average normalized PSTHs are shown to the left and below the JPSTH. The JPSTH is symmetric due to a randomization procedure, which eliminates a positioning bias of cells on either the x- or y-axis. JPSTH diagonal (black line) was computed by averaging ±25ms around the main diagonal. **(B)** Average main JPSTH diagonal ±SEM of all task-modulated cell pairs. Black horizontal bar indicates significance (t-test, $p<0.05$). Inset: Average main JPSTH diagonal after removal of the outer 20% of cell pairs determined by the average strength of correlated firing. **(C)** Correlation of JPSTH diagonal coefficients and Pearson correlation coefficients. The correlation was computed for the average coefficients in the 300-600ms post attention cue onset. Histograms represent distribution of values. R^2 and p-value result from the regression analysis. **(D)** Average Pearson correlation coefficient ± SEM over all task-modulated cell pairs. P-value represents outcome of a t-test comparing the average coefficient values in the 300-600ms post attention cue onset (grey zone) against zero. **(E)** Same conventions as in **A**, but the JPSTH was created with a 10ms smoothing window. **(F)** Same conventions as in **B**, but created using a 10ms smoothing window.

Coincidence lines were then baseline-corrected by subtracting the average correlated spiking activity within 200 ms before the attention cue onset. Significance along the diagonals, averaged over all cell pairs or over cell pairs of a given area combination, was assessed using a t-test ($p \leq 0.05$). Task-modulation of cell pairs was defined as a significant change (t-test, $p \leq 0.05$) in average firing rate in both cells over the 800 ms after the attention cue presentation in comparison to the 200 ms before the attention cue. When computing the average JPSTH, to prevent the arbitrary positioning of each cell on either the x- or y-axis from confounding the overall JPSTH, we randomly placed each cell in a pair on either the x- or y-axis 200 times and averaged across randomizations. This procedure led to a symmetric average JPSTH (Figure 2.2A). When computing the average percent change of correlation before and after the attention cue, for each cell pair the average correlation in the 200 ms before cue onset was compared to the 300-600 ms time window after cue onset.

For analyzing the firing rate dependence of synchronized firing, we normalized the PSTHs of each cell by subtracting the baseline period in the 200ms before cue onset. We then averaged the normalized firing rate of both cells in a given pair for the time window 0 - 800 ms after the attention cue onset and used a regression analysis ($p \leq 0.05$) to determine if a significant correlation existed between the average firing rate and strength of synchronous firing in each cell pair. When analyzing the distance dependence of synchronous firing, we measured distance between two cells in two different ways. First, we measured the

physical distance between cells based on the x-y-z 3-D coordinates of the electrode tip they were recorded from, and second, we measured cortical grey-matter distance between electrode tips based on the 2-D coordinates on a flattened representation of the cortical sheet (see Kaping et al., 2011). For both measures, correlation between cell distance and firing correlations in the time window 0 - 800 ms after cue onset was evaluated using a regression analysis ($p \leq 0.05$).

When analyzing the temporal characteristics of neuronal firing correlation, we adjusted the Gaussian kernel widths with which the PSTHs were computed to 12.5, 25, 50 (default), 100 and 200 ms and repeated the analysis. These kernel widths were defined as time windows of 25, 50, 100, 200, and 400 ms, and will be referred to as such from now on. Based on the temporal evolution of the basic firing correlations, we compared the strength of the effect at different smoothing windows in the first (0-300) and second 300 ms window (300-600) after attention cue onset using varying Gaussian kernel widths. Differences were assessed using a t-test and one-way ANOVA ($p \leq 0.05$).

Anatomical specificity

To test for anatomical specificity of firing correlations, we separated cell pairs into four *within-area* groups (ACC-ACC, latPFC-latPFC, dPFC-dPFC, vmPFC-vmPFC), and six *between-area* groups (ACC-dPFC, ACC-latPFC, ACC-vmPFC, dPFC-latPFC, dPFC-vmPFC, latPFC-vmPFC), based on the subnetwork area division scheme recently proposed by Saleem et al., (2014) (Figure 2.1B). To determine significant firing correlations exceeding zero correlations between specific area combinations, all JPSTH coincidence lines from cell pairs within a specific area combination were averaged. Based on the temporal development of the increased synchronous firing across area combinations, comparisons were made within the 300-600ms time window after attention cue onset. Significant differences from zero were determined with a t-test ($p \leq 0.05$).

Asymmetry of spiketrain correlations

We calculated an asymmetry index (see Paz et al., 2007) to test for systematic time lags underlying the firing correlations of the spiking activity of two neurons in separate areas (see Figure 2.4A). This was done by separately adding all bins above (a) and below (b) the coincidence line of the JPSTH for the time window 0 - 800 ms after cue onset and computing a normalized difference (a-b)/(a+b) for each JPSTH. A significant difference from zero of the average asymmetry index for each area combination was tested using a Wilcoxon signed rank test ($p \leq 0.05$) and implied that spikes from one area were systematically leading/lagging spiking activities in another area.

Spatial attention analysis

We determined the contra- and ipsilateral preference of each cell using their relative firing rate modulation after attention cue onset. If a cell showed on average a higher firing rate within the 800 ms after the attention cue onset when attention was shifted to the contralateral as opposed to the ipsilateral side, this cell was labeled as a contralateral preferring cell and vice versa. Accordingly, cell pairs were then labeled as either contra-contra, ipsi-ipsi, or contra-ipsi preferring pairs. Average spiketrain correlations between these three groups were compared by averaging the JPSTH coincidence lines of all cell pairs within a group and comparing these across the three groups using an ANOVA and posthoc comparisons (1-way ANOVA, $p \leq 0.05$; Tukey-Kramer test, $p \leq 0.05$).

We determined for every cell pair whether their spiketrain correlations contained spatial attention information using a randomization analysis. For every cell pair, trials were randomly labeled as attend-right or attend-left, and the resulting average JPSTH was created 300 times. The average randomized JPSTH coincident line was then compared against the original coincident line. A cell pair was considered to show significant spatial attention effects when, within 0 - 800 ms after the attention cue onset, firing correlations were for a minimum of 150 ms significantly different from the randomized average coincident line (|average diagonal|>1.96*Standard deviation of the randomization statistic). We then computed the fractions of significant cell pairs for each area combination separately and compared them

for *within-area* combinations and for *between-area* combinations using a chi-square statistic ($p \leq 0.05$). We additionally tested whether each fraction individually was significantly different from what would be expected by chance (Binomial statistics, $p \leq 0.05$).

Classification of cell types according to potential waveform characteristics

For the set of highly isolated neurons (n = 404) in our sample, we aligned, normalized and averaged all action potentials (APs) (For details of these procedures, see Ardid et al. (2015) and the freely available online repository and matlab documentation at http://attentionlab.ca/doku.php?id=ap-waveform-analysis). Each neuronal waveform was fitted with cubic interpolation from an original precision of 25μs to 2.5μs. On the resultant waveform, we analyzed two measures (Figure 2.6B): the peak-to-trough duration and the time for repolarization. The time for repolarization was defined as the time at which the waveform amplitude decayed 25% from its peak value. These two measures were highly correlated ($r = 0.68$, $p < 0.001$, Pearson correlation). We computed the Principal Component Analysis and used the first component (explaining 84.5 % of the total variance), as it allowed for better discrimination between narrow and broad spiking neurons, compared to any of the two measures alone. Results from the calibrated Hartigan Dip Test discarded unimodality for the first PCA component ($p < 0.01$) and for the peak to trough duration ($p < 0.05$) but not for the duration of 25% repolarization ($p > 0.05$). In addition, we applied Akaike's and Bayesian information criteria for the two- versus one- Gaussian model to determine whether using extra parameters in the two-Gaussian model is justified. In both cases, the information criteria decreased (from -669.6 to -808.9 and from -661.7 to -788.9, respectively), confirming that the two-Gaussian model is better. We then used the two-Gaussian model and defined two cutoffs that divided neurons into three groups. The first cutoff was defined as the point at which the likelihood to be a narrow spiking cell was 10 times larger than a broad spiking cell. Similarly, the second cutoff was defined as the point at which the likelihood to be a broad spiking cell was 10 times larger than a narrow spiking cell. Thus, 95% of neurons (n = 384) were reliably classified: neurons at the left side of the first cutoff

were reliably classified as narrow spiking neurons (20%, n = 79), neurons at the right side of the second cutoff were reliably classified as broad spiking neurons (75%, n = 305). The remaining neurons were labeled as 'fuzzy' neurons as they fell in between the two cutoffs and were not reliably classified (5%, n = 20), see Figure 2.6). All isolated waveforms that were not separable from the noise in a clean way were defined as multi unit activity (MUA) and were likely composed of narrow as well as broad spiking cells.

Pearson correlation analysis

In order to verify our main results with a somewhat simpler approach, we computed the Pearson correlation coefficients on the spike density functions of all our cell pairs. The Pearson correlation coefficient is defined as follows:

$$r_{i,j}(t) = \frac{\Sigma(S_i(t) - \overline{S}_i(t))(S_j(t) - \overline{S}_j(t))}{\sqrt{\Sigma(S_i(t) - \overline{S}_i(t))^2 \Sigma(S_j(t) - \overline{S}_j(t))^2}} \quad \text{(eq. 3)}$$

where $S_i(t)$ represents the activity of neuron *i* at time t, $\overline{S}_i(t)$ represents the average activity of neuron i at time t over all trials. Coefficients were computed in 5ms bins for the time window -0.2 - 0.8 ms following attention cue onset. For averaging, Pearson correlation coefficients were first normalized by Fisher z-transformation and then in line with the JPSTH analysis baseline corrected by subtracting the average activity in the 200ms before attention cue onset. Significant difference from zero for the average normalized Pearson correlation coefficient was determined in the time window 300-600 ms post attention cue onset with a t-test ($p \leq 0.05$) (Figure 2.2D). To qualitatively compare the JPSTH with the Pearson correlation method, we computed the average Pearson correlation coefficient over the 300-600ms time window of interest for each cell pair and compared it against the JPSTH diagonal averaged over the same time window. Significant correlation between the two was determined with a regression analysis ($p \leq 0.05$) (Figure 2.2C). Additionally, we computed the Spearman correlation coefficients and analyzed them

the equivalent way to the Pearson correlation coefficients. Results were qualitatively similar (data not shown).

Analysis of reward outcome expectations on spiketrain correlations

Cued target stimuli were associated with a high- or low- reward outcome, which allowed testing for possible influences of reward outcome expectation on spiketrain correlations (for details, Kaping et al., 2011). For this analysis, trials were split into those with high- and low- reward outcome associated with the target stimulus for each cell pair before computing the JPSTHs. We then compared the average diagonals across cell pairs for low- and high-reward outcome trials within the 300 - 600ms after the attention cue onset (t-test, $p \leq 0.05$). This was done for the task-modulated cell selection as well as over all cell pairs.

Testing the relation of spiketrain correlations and burst firing

To investigate the relation between spiketrain correlations and the occurrence of burst firing, which was previously reported to be correlated with LFP activity (Womelsdorf et al., 2014a), we extracted all neuron pairs in which either both neurons or one neuron were recorded as single units (SUA). Burst events were defined as 2 or more spikes occurring within 5ms (for details *see* Womelsdorf et al., 2014a). The proportion of burst to non-burst events was calculated for the 300-600ms post attention cue onset period for neurons identified as SUA only. If both neurons in a pair were SUA, the proportion of burst to non-burst events was averaged over the two neurons. We then compared the proportion of burst to non-burst events for each neuron pair with the strength of spiketrain correlations (Regression analysis, $p \leq 0.05$).

Comparison of spiketrain correlations and phase-amplitude LFP-LFP correlations

In order to compare the spiketrain correlations with cross-frequency correlations of LFP activity that was previously reported on the same dataset (Voloh et al., 2015), we first extracted LFP pairs for those

recording pairs that also contained neuronal firing activity and were thus used in the spiketrain correlation analysis. We then followed procedures identical to those outlined in Voloh et al., 2015. In brief, we bandpass filtered the raw LFP signals with a 4th order, two pass Butterworth filter followed by Hilbert transforming them. We restricted this analysis to those frequencies that were shown to be related to attention in the previous report, extracting the (theta) phase of the 7±2.3Hz as [4.7 9.3] Hz, and the (gamma) amplitude envelopes of the 40 ±13.3Hz as [26.7 53.3] Hz frequency bands for all correctly performed trials in the [-0.5 0] sec. and [0 0.5] sec. time window before and after attention cue onset. We quantified cross frequency correlation between the theta phase and the gamma amplitude using Tort's Modulation Index (Tort et al., 2010). Phase amplitude correlation was recorded exclusively between electrodes recorded from different electrodes and only for LFP pairs recorded from electrodes that also provided neuronal firing signals used in the main analysis of spiketrain correlations. Theta-gamma correlations and spiketrain correlations were compared for matching electrode pairs using Pearson correlation coefficients after removal of outliers which were defined as values >5 STD of the z-transformed data.

2.4 Results

We recorded from 1151 neurons in macaque anterior cingulate and lateral prefrontal cortex (ACC/PFC) while animals performed a cued spatial attention task (Figure 2.1A), described in detail elsewhere (Kaping et al., 2011). Results from qualitatively different analyses of this dataset have previously been reported (Kaping et al., 2011; Shen et al., 2014; Womelsdorf et al., 2014a, Voloh et al., 2015). During task performance, we collected 2131 pairs of cells recorded simultaneously from different electrodes. We did not include same-electrode pairs in this analysis, and restricted analysis to correctly performed trials with equal sensory stimulation conditions during the entire analysis time windows (see Methods).

2.4.1 Behavioral performance

The two animals exhibited similar performance, with 71% (Monkey M) and 63% (Monkey R) correctly completed trials, respectively. Error responses were also distributed similarly across the two monkeys, whereby they consisted of fixation breaks occurring before the rotation of the target stimulus (Monkey M: 25%, Monkey R: 20%), incorrect responses defined as saccadic responses to the wrong response target (Monkey M: 1%, Monkey R: 3%), late responses initiated only after the response time window (Monkey M: 0%, Monkey R: 1%), no responses (Monkey M: 7%, Monkey R: 1%), and hold fixation on target errors (Monkey M: 2%, Monkey R: 6%).

2.4.2 Spiketrain correlations increase when selective attention is deployed

To investigate potential neural interactions between cells across the prefrontal and cingulate cortices, we analyzed coincident firing (henceforth named correlated firing) by computing joint peristimulus time histograms (JPSTHs, Figure 2.1C, D), that allow measurements of dynamic crosscorrelations of spiketrains of distant neurons (Aertsen et al., 1989). JPSTHs were computed using 100ms smoothing time windows (see Materials and Methods), and increases and decreases of firing correlations between cells were assessed via the coincidence line of each JPSTH and validated by calculating Pearson correlations of the spiketrains in the same analysis window (Figure 2.1C**,** see Materials and Methods). In order to test whether neurons dynamically synchronize their firing when attention is shifted, JPSTHs were aligned to the onset of the attention cue (Figure 2.1A). When we averaged the coincidence lines of all cell pairs that showed significant task modulation (see Materials and Methods) we found that cell pairs across the prefrontal and anterior cingulate cortices showed a transient increase in firing correlations during the 280-550 ms after the attention cue onset (Figure 2.2A, B; t-test, $p \leq 0.05$). This effect remained evident when considering all cell pairs irrespective of the cells' task modulation (data not shown). Additionally, this effect disappeared when considering only non-task modulated neuron pairs (n=120; t-test, p=0.194; data not shown). The increase in firing correlations following the attention cue was not due to transient

increases in average firing rates, as can be seen in the average PSTHs (Figure 2.2A, see also example JPSTHs in Figure 2.1C, D). Previous studies using the JPSTH controlled for the influence of overall firing rate magnitudes to affect correlations by computing surrogate data with the same average firing rates but where specific spike timing information was random (e.g. Cohen et al., 2010). To exclude the possibility that our main effects are due to spurious JPSTH correlations driven purely by mean firing rate, we correlated the average firing rate of neurons in a given pair with the average JPSTH diagonal, and found that overall firing rates did not explain significant spiketrain correlations (Figure 2.3A). To test whether the main effect depended on cell pairs with particularly high or low correlation values, we calculated the effect after removing cell pairs with the strongest 20% correlate values. As shown in Figure 2.2 (inset), the main effect persistent for this reduced population of cell pairs. As a control analysis, we computed the Pearson correlation coefficients for each cell pair. The average Pearson correlation coefficients for all task modulated cell pairs was significantly increased in the 300 - 600ms post attention cue onset (Figure 2.2D; t-test, $p = 0.028$). This effect also remained when considering all cell pairs independent of task modulation (data not shown, t-test, $p = 0.015$). Normalized Pearson correlation coefficients were highly correlated with the normalized JPSTH diagonal averaged over the time window 300 - 600 ms post attention cue onset (Figure 2.2C; Regression analysis, $p = 6.82*10^{-119}$). Spearman correlation coefficients were similarly correlated with the JPSTH diagonal values (data not shown, Regression analysis, $p = 3.89*10^{-80}$). We next tested whether firing correlations varied systematically as a function of the physical (3D) anatomical distance between cell pairs. We found that firing correlations tended to decrease with distance (regression analysis, $p = 0.084$), but that this trend was driven by neuron pairs recorded from nearby electrodes approximately 500µm apart (Figure 2.3B). These *nearby* recordings were made with 125µm thick tungsten electrodes glued to each other with a 500µm tip-to-tip distance and lowered within the same guide tube. Beyond the immediate ~500µm distance there were no simple distance dependencies of firing correlations evident ($p = 0.86$).

To provide an understanding of the overall effect size and distribution, single correlation values in the

JPSTHs of all cell pairs fell within the range of -3.82 and 4.44; whereby 90% of the data points lay within -0.22 and 0.23. The average increase in spiketrain correlation ranged from -0.01 to 0.02 during the post-cue period (see Figure 2.2B), with the average percent change from prior to attention cue onset to the 300 - 600ms post attention cue onset reaching 106%.

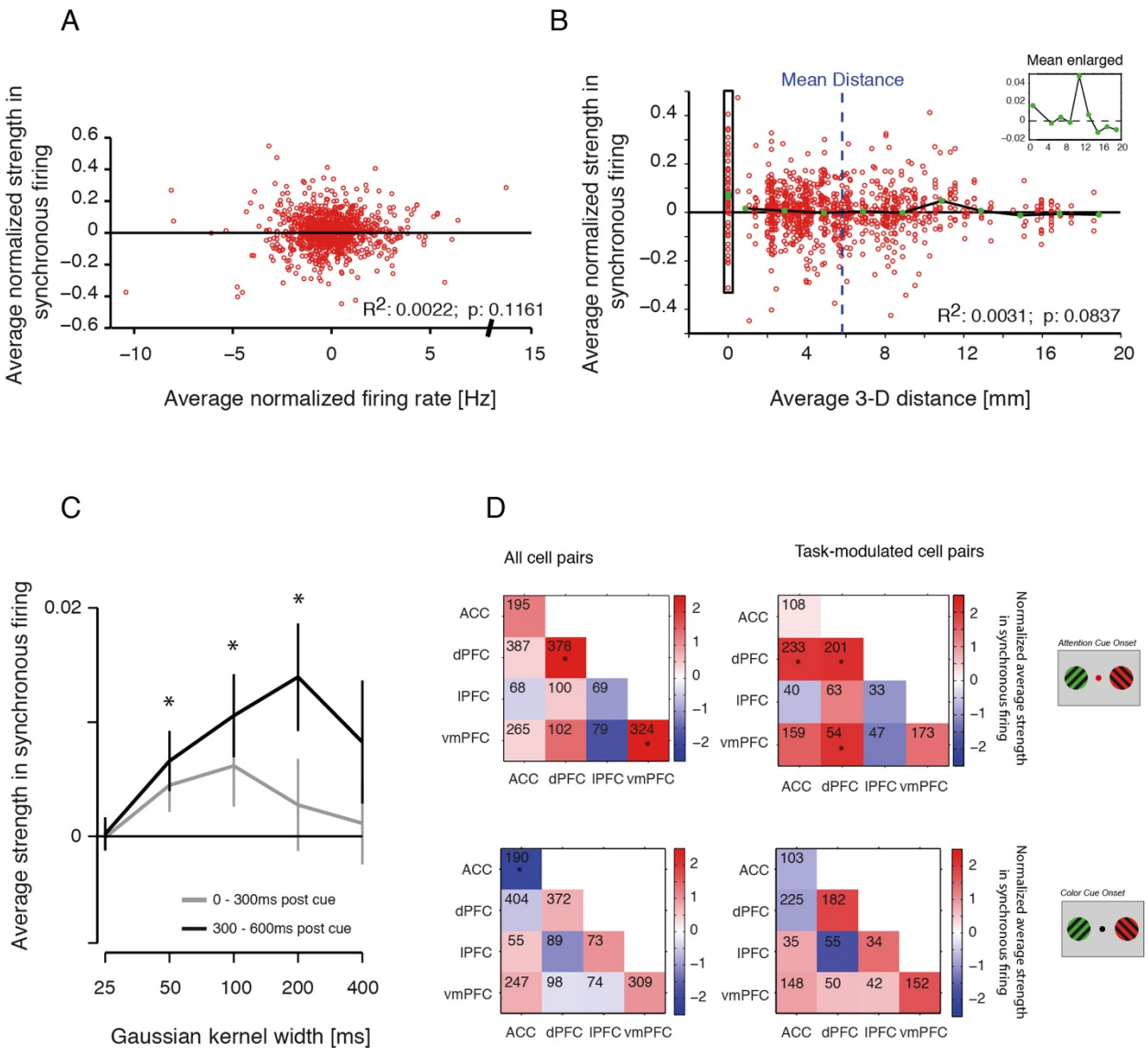

Figure 2.3 Firing Correlations and their temporal and spatial specificity.

(A) Correlation of the average normalized firing rate (*y-axis*) plotted against the average normalized firing correlation of cell pairs within 800 ms after the attention cue onset. The average firing correlation corresponds to the coincident firing ±25ms in Figure 2.2. R^2 and p-value indicate the result of the

regression analysis. **(B)** Correlation of the physical 3D distance between cells and the strength of their correlated firing. Every data point represents one task-modulated cell pair plotted as the distance between the two cells against their firing correlation strength. Green dots indicate the average strength of synchronous firing for every 2mm distance bin. The vertical rectangle on the left highlights the data points from cell pairs recorded from electrode tips that were ~500µm apart from each other (see Materials and Methods). R and p-value indicate regression analysis results. The inset shows a larger view of the average effect. The blue dotted line represents the mean distance across all task-modulated cell pairs. Please note that quantitatively similar results were obtained when 2-D distances (based on flattened map of the cortical sheet, see Kaping et al., 2011 for details) were used as distance estimates. **(C)** The effect of the JPSTH smoothing kernel width on the average firing correlation strength. The window after the attention cue onset was split into two representative time windows - 0 - 300ms after cue onset and 300-600 ms after cue onset, shown by the grey and black line, respectively. The significantly increased firing correlation (Figure 2.2A) in the late 300ms after cue onset was evident for smoothing kernel widths ranging from 50-200ms, but absent when smaller or larger smoothing was applied (t-test, $p<0.05$). **(D)** Adjacency matrices showing the firing correlations within 300-600 ms after the attention cue onset for all area combinations. Numbers within squares indicate the number of cell pairs for that respective area combination. Black asterisks indicate firing correlations significantly different from zero. Top row shows alignment to attention cue onset, bottom row shows alignment to color cue onset. Left column shows results for all cell pairs, right column shows results for task-modulated cell pairs.

2.4.3 Temporal specificity of spiketrain correlations

We next tested whether firing correlations in ACC/PFC were specific to a time scale of 100 ms, or whether firing correlations were evident also with shorter or longer time windows. To test for this temporal specificity of functional correlations we calculated for each neuron pair the coincidence line of the JPSTH computed over time windows of 25, 50, 100, 200 and 400 ms. To assess the influence of the smoothing, we extracted the average spiketrain correlations within the first 300 and second 300 ms after attention cue onset (Figure 2.3C). Within the second 300 ms, spiketrain correlations calculated with smoothing kernels of 50, 100 and 200 ms were significantly larger than zero (t-test, $p \leq 0.05$), while the effect disappeared at 25ms and 400ms smoothing width. A one-way ANOVA confirmed that spiketrain correlations within the first and the second 300ms after attention cue onset were significantly different from each other (ANOVA, $p \leq 0.05$).

To show nearly unsmoothed JPSTHs and to further demonstrate the temporal aspect of the spiketrain correlations, we also computed JPSTHs with 10ms time windows (5ms Gaussian kernel, see Material and Methods). The single JPSTHs (Figure 2.1D) and the average JPSTH and diagonal (Figure 2.2E, F) show

that spiketrain correlations with this smoothing factor are highly variable and don't show the same consistency as when larger smoothing factors are applied.

2.4.4 Anatomical specificity of spiketrain correlations

To quantify whether functional spiketrain correlations were specific to neuron pairs from particular cortical subfields in ACC/PFC, we localized their anatomical location to one of four subfields as delineated by Saleem et al. (2014, see Figure 2.1B) and compared spiketrain correlations for each area combination separately (*see* Materials and Methods). We sampled sufficient pairs between four subfields with neurons located in either of three prefrontal cortex areas or in anterior cingulate cortex: lateral prefrontal cortex (area 46), dorsal prefrontal cortex (areas 8, 9), as well as ventromedial prefrontal cortex (area 32), and the anterior cingulate cortex (area 24). Significant spiketrain correlations were found for subsets of area combinations only. Significant spiketrain correlations were evident for *between-area* cell pairs in ACC and dPFC (t-test, $p = 0.0405$), and *between-area* cell pairs in vmPFC and dPFC (t-test, $p = 0.0395$) (Figure 2.3D). These two *between-area* effects were only found in cell pairs in which both cells showed significant task-modulation (see Methods). Increased spiketrain correlations 300 - 600ms after attention cue were also found *within-areas*, for cell pairs in which both cells were located *within* dPFC (t-test, $p = 0.0036$) (Figure 2.3D), and for cell pairs in which both cells were located *within* vmPFC (t-test, $p = 0.0012$) (Figure 2.3D). These effects were not dependent on task-modulation and became more pronounced when averaging over all cell pairs in the case of *within-area* dPFC cell pairs. These results show that spiketrain correlations were largely restricted to specific cortical subfields at specific time windows in ACC/PFC.

2.4.5 Correlated firing is maximal during attention shift

We next were interested in whether spiketrain correlations within ACC/PFC are functionally tied to

covert stimulus selection during the attention shift or whether they may characterize spiketrain interactions even prior to the attention cue onset. We thus tested whether spiketrain correlations were evident with all data aligned to the color onset (Figure 2.1A). The only difference visually between the color onset and attention cue onset is the color of the attention cue, which instructs the attention shift. When we performed this analysis for task-modulated cell pairs, none of the area combinations tested showed significant firing correlations in the 200-500ms after the color cue onset (Figure 2.3D, t-test, $p > 0.05$). When all cell pairs were considered, neurons *within* ACC showed significant anti-correlation, an effect that was not observed within the attention cue period. Similar results were obtained when the analysis window for the color cue period contained the 0 - 300ms post color cue onset period, or the 300ms immediately prior to attention cue onset. This suggests that the increase in correlated firing between cells in ACC/PFC is specific to processes of stimulus selection and attention shifts. The transient nature of this effect is further confirmed by analyses that show no increase in spiketrain correlations prior to the response event (data not shown).

2.4.6 Asymmetry of spiketrain correlations

To investigate whether the firing of one neuron systematically leads or lags the spiking activity of another neuron we computed an asymmetry index of the JPSTH (Paz et al., 2007). The asymmetry index (AI) quantifies whether spiking coincidences above versus below the coincidence line of the JPSTHs systematically differ (Figure 2.4A). In the time window encompassing 800 ms following the attention cue, the AI was significantly different from zero for neuron pairs from ACC - dPFC indicating that spikes from neurons in dPFC tended to precede spiking activity from neurons in ACC (Figure 2.4B, C; Wilcoxon signed rank test, $p = 0.0114$). This effect was reliably observed for smoothing time windows 25, 50, 100 and 200 ms in task-modulated cell pairs (Wilcoxon ranksum test; $p = 0.0018$, $p = 0.0036$, $p = 0.0114$, $p = 0.0227$, respectively) and for all smoothing time windows when all cell pairs were taken into

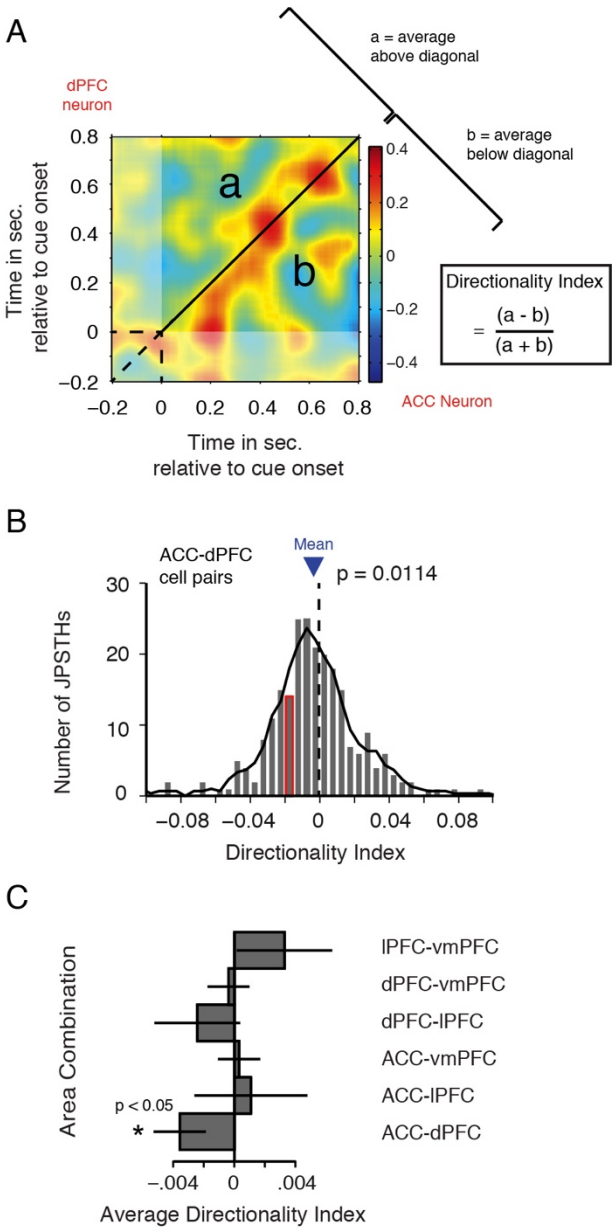

Figure 2.4 Asymmetry of Correlated Firing.

(A) Example JPSTH of a cell pair with one cell in ACC (*x-axis*) and the other in dorsal PFC (*y-axis*) illustrating the calculation of the asymmetry index (AI). The AI was computed over the time window 0-800ms post attention cue onset with the smoothing window of 100ms for all task-modulated cell pairs. An average value above the diagonal and below the diagonal is computed and the AI is obtained using the formula to the right of the JPSTH. **(B)** Histogram of the distribution of AI's across all ACC-dPFC cell pairs. The mean AI (-0.0187) is indicated by the blue triangle. The example pair from **A** falls into the bar highlighted with a red border. The p-value indicates that the distribution of DIs falls significantly below zero (Wilcoxon rank sum test). **(C)** Average directionality indices for every area combination. Error bars represent SEM. Black asterisks indicate significance (Wilcoxon rank sum test, $p<0.05$).

account (Wilcoxon ranksum test; p = 0.0004, p = 0.0003, p = 0.0017, p = 0.0048, p = 0.0070, respectively). When considering only the fraction of the JPSTH within the time window 300 - 600 ms, this prior reliably significant effect was observed only for the 400ms (p = 0.0128) and as a trend for the 200ms (p = 0.0803) smoothing window in task-modulated cell pairs, and for 200ms (p = 0.0475) and 400ms (p = 0.0049) smoothing when all cell pairs were taken into account.

2.4.7 Spiketrain correlations of cell pairs with similar spatial attention preferences

To investigate whether synchronous firing contains relevant information for the attention shift, we asked whether spiketrain correlations depended on the relative tuning preferences of individual neurons to increase their firing for contra- or for ipsilateral attention shifts. To this end, we classified neurons into those with firing increases for contra- or ipsilateral attention shifts during the attentional shift epoch and compared the sign and strength of spiketrain correlations among pairs with similar and dissimilar spatial attention preferences. Overall, we obtained three groups with 399 ipsi-preferring cell pairs, 598 contra-preferring cell pairs, and 918 mixed preference cell pairs. We found that, on average cell pairs in which both cells preferred an attention shift to the contralateral side were significantly more likely to show spiketrain correlations within 0 - 800ms following the attention cue, than pairs in which both cells preferred an attention shift to the ipsilateral side (1-way ANOVA, $p < 0.05$; Tukey-Kramer test, $p < 0.05$; Figure 2.5A). Ipsilateral-preferring cell pairs show a trend for anti-correlated firing. The previous findings suggest that spiketrain correlations carry information about the direction of the attention shift and may be functionally instrumental for attention. To test further whether this functionally meaningful correlation can be identified at the level of single cell pairs, we determined for every cell pair whether its firing correlation was significantly stronger for either contra- or ipsilateral attention shifts as tested for significance using a randomized distribution (±2SD). We then computed the fraction of significant cell pairs for each area combination separately (Figure 2.5Bi-iii). We found that cell pairs between ACC and

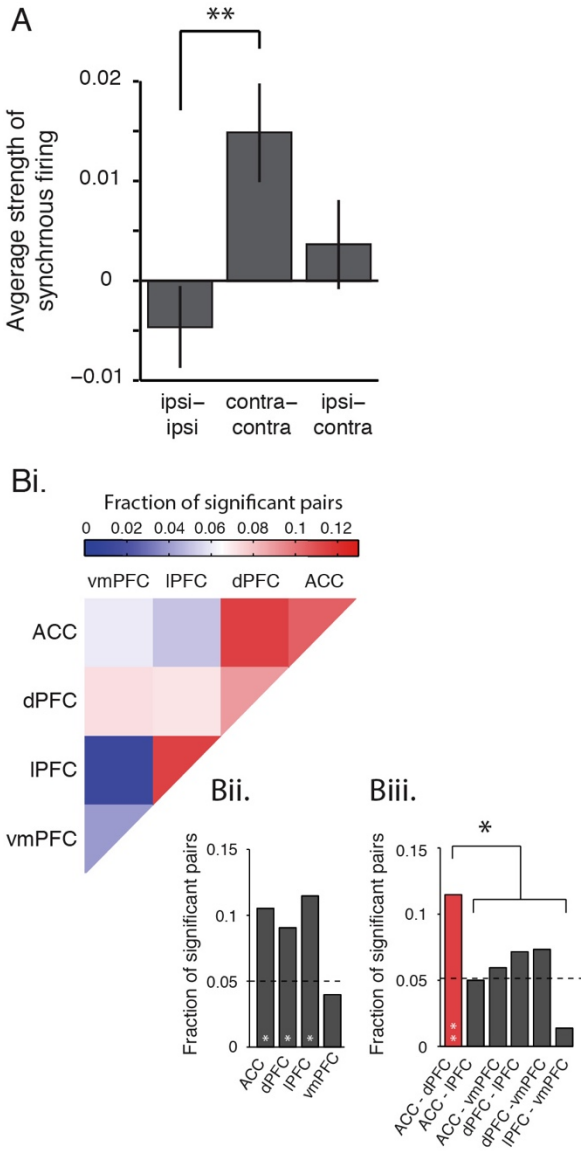

Figure 2.5 Firing Correlations carry spatial attention information.

(A) Average correlation for cell pairs in which either both cells preferred ipsi-lateral attention shifts, both cells preferred contra-lateral attention shifts, or cells preferred opposite attention shifts (ipsi-contra). Error bars represent SEM. The black asterisk indicates significant posthoc comparison with a Tukey-Kramer test ($p \leq 0.01$). **(B)** Fraction of cell pairs that have significantly differing firing correlations for ipsi- versus contra-lateral attention shifts in the 800ms following the attention cue onset, separated by area combination. **i.** Adjacency matrix highlighting the fraction of significant ($p < 0.05$, randomization statistic) cell pairs for each area combination. **ii.** and **iii.** show for visualization purposes the same as **i.**, separated for within- and between-area combinations, respectively. Black asterisks indicate significant difference from all other area combinations (chi-square test). White asterisks inside bars indicate significant difference from what would be expected by chance (Binomial test, *$p<0.05$, **$p<0.001$).

dPFC, as well as within ACC, dPFC, and lPFC had a significantly larger fraction of significant pairs than would be expected by chance (Binomial test, ACC-dPFC: $p=1.2*10^{-6}$; ACC: p=0.003; dPFC: p=0.001; lPFC: p=0.023). We additionally found that ACC and dPFC pairs contained a significantly larger fraction of significant cell pairs compared to the average of all other *between-area* pairs (Figure 2.5Biii; chi-square, $p<0.05$). In other words, cell pairs in ACC and dPFC showed an increased selectivity for attention shifts to one of two possible spatial targets, in comparison to neurons from other *between-area* combinations.

2.4.8 Cell-type specificity of spiketrain correlations

We next asked whether positive or negative spiketrain correlations operating on the 100 ms time scale are differentially subserved by excitatory and inhibitory cell classes, respectively. Typically, the influence of excitation and inhibition on postsynaptic target cells are measured at shorter millisecond time scales (Reid and Alonso, 1995). In contrast to such effective connectivity measures our analysis attempts to identify whether firing correlations can be traced back to slower (100ms) interactions among excitatory neurons, and whether anti-correlated firing can be traced back to the influence of putative interneurons that may actively suppress their target neurons. To answer this question, we selected from the total population of recorded neurons those 404 neurons that were unambiguously recorded as single isolated neurons (see Materials and Methods, and Shen et al., 2014). Figure 2.6A illustrates that the neuron population separated into narrow spiking (NS) cells and broad spiking (BS) cells that correspond to predominantly inhibitory interneurons (NS cells) and predominantly pyramidal cells (BS cells). Using the duration and time to repolarization of the action potential waveforms revealed a significantly bimodal distribution ($p<0.001$, calibrated Hartigan's dip test for unimodality) with 21.6% (n=86) NS cells, 74.4% (n=296) BS cells, Figure 2.6B, C) (see Methods and Ardid et al., 2015). These proportions are similar to the 13-25% NS cells previously reported for the lateral prefrontal cortex (Constantinidis and Goldman-Rakic, 2002; Diester and Nieder, 2008; Johnston et al., 2009). The split into two cell classes allowed quantifying the

sign and strength of spiketrain correlations separately for NS and BS cells, and allowed testing how isolated BS and NS cell firing related to the multiunit activity MUA that is composed of mixtures of BS and NS cells (Figure 2.6D). Figure 2.6E illustrates that significant spiketrain correlations are evident across all pairs of BS cells and MUA's ($p = 0.0002$, Wilcoxon rank sum test, overall 631 pairs with at least one BS cell). In contrast, neuron pairs with one identified NS cell and MUA showed no significant correlation, but rather showed a trend for anti-correlated firing (Wilcoxon rank sum test, $p=0.066$, $n=162$ pairs) (Figure 2.6E). Neuron pairs with an identified BS cell and NS cells did not show significant spiketrain correlations (Wilcoxon rank sum test, $p=0.68$, $n=62$ pairs) (Figure 2.6E).

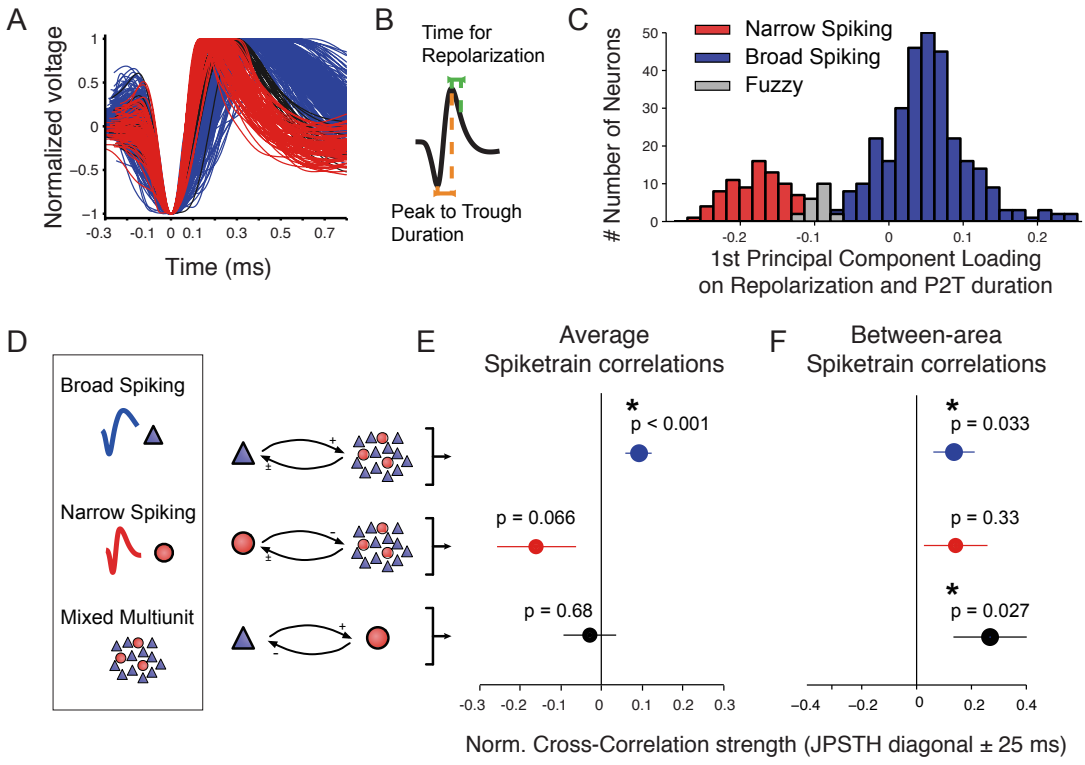

Figure 2.6 Cell type specificity of firing correlations in ACC/PFC.

(**A**) Normalized action potential (AP) waveforms of single cells. Red and blue colors highlight narrow spiking (NS) and broad spiking (BS) cells. Black waveforms mark cells that were not classified (fuzzy category). (**B**) Illustration of the two features used to characterize cells AP's. (**C**) Bimodal distribution of cells according to the combination of the two action potential features, showing clean separation of NS

and BS cells (see Materials and Methods). (**D**) Illustration of combinations of BS, NS and multiunit recordings tested for consistent spiketrain correlations (**E**) Normalized cross-correlation, indexed as ±25ms average JPSTH diagonal (*x-axis*) for three different types of cell combinations: The blue mark denotes spiketrain correlations between BS cells and multiunits. The red mark shows (anti-) correlation of NS cell spiketrains with multiunits. The black mark denotes spiketrain correlations of NS with BS cell activity. P-values show significance of the differences to zero correlation (*Wilcoxon rank sum test*). (**F**) Same format as E but restricting the averaging of spiketrain correlation to pairs of neuron activity from different brain areas. Error bars in E and F denote standard errors of the mean. Firing correlations were indexed as the average coincidence line (±25ms) of the JPSTH during 0.3 - 0.6 sec. following attention cue onset.

The previous analysis suggests that functional firing rate correlations in ACC/PFC can be traced back to single, isolated BS cells showing positive correlations with MUA activity. To test whether the sensitivity of our measure is sufficient to resolve correlations of single neurons located in different cortical fields, we performed the analysis only on *between-area* pairs. Intriguingly, we found that the spiketrains of BS cells remained significantly correlated with MUA activity even when BS and MUA were recorded from different cortical fields (Figure 2.6F, Wilcoxon rank sum test, $p<0.05$, n=342 pairs (BS-Multiunit)). In addition, we found that the firing of isolated BS cells in one cortical field and the activity of NS cells in another cortical field were on average positively correlated (Figure 2.6F, Wilcoxon rank sum test, $p<0.05$, n=34 pairs (BS-NS)). We did not find evidence for NS cells engaging significantly in anticorrelated firing with BS, MUA, or NS cells within or between specific cortical fields (data not shown).

2.4.9 Spiketrain correlations are not related to expected reward magnitude

To assess whether the magnitude of the reward outcome associated with the target stimulus affected spiketrain correlations, we compared trials with targets of high versus low reward outcome. We found no difference in the average main diagonal within the 300 - 600ms post attention cue onset between the high and low outcome condition for either task-modulated (t-test, $p = 0.481$) or all cell combinations (t-test, $p = 0.274$) (data not shown). This suggests that reward outcome expectation was not a driving factor for the spiketrain correlations following the attention cue onset.

2.4.10 Spiketrain correlations show no apparent relation to the occurrence of burst firing

We showed previously that burst firing synchronized with distant local field potentials at narrow frequency bands (see Womelsdorf et al., 2014a for details), raising the possibility that burst firing events may contribute to the correlation of firing rates between areas. We therefore tested whether the spiketrain correlations observed here were related to the occurrence of burst firing. To this end we extracted neuron pairs where either both or one of the neurons were recorded as single unit activity (SUA) and extracted for each pair the proportion of burst to non-burst events. Larger proportions of burst firing would be expected to positively correlate with the strength of JPSTH correlations if burst firing plays a prominent role underlying spiketrain correlations. However, we found that the proportion of burst firing and the strength of spiketrain correlations were not correlated for either neuron pairs in which both cells were recorded as SUA (n = 72; Regression analysis, $R^2 = 0.009$, $p = 0.426$) or for neuron pairs in which only one cell was recorded as SUA (n = 893; Regression analysis, $R^2 = 0.0004$, $p = 0.541$). This finding suggests that the spiketrain correlations were not apparently linked to mechanisms underlying burst firing.

2.4.11 Spiketrain correlations show no apparent relation to LFP theta-gamma phase amplitude correlations

A recent study using the local field potentials of the same dataset used here found inter-areal interactions between the ACC and the lateral PFC in local field potential activity (Voloh et al., 2015). This study reported that attention cues are followed by (5-10 Hz) theta frequency phase modulations of the LFPs correlated with the amplitude of 35-55 Hz gamma frequency amplitude modulation of LFPs between ACC and prefrontal cortical areas (Voloh et al., 2015). We were thus interested in whether this attention-related phase amplitude correlation of LFP activity is related to the inter-areal spiketrain correlations. To test this, we extracted the LFP of those electrode pairs that contributed neurons to the spiketrain correlation analysis. We then calculated for each LFP pair the 5-10 Hz (theta) phase to 35-55 Hz (gamma) amplitude variation in the post attention cue interval that was previously shown to have significant

attention effects at the overall population level and across a large fraction of single LFP pairs (see Materials and Methods, Voloh et al., 2015). We then calculated correlation coefficients of the spiketrain correlations in the 300-600ms, and the strength of theta-gamma correlation in the post-cue period calculated as modulation index (Tort et al., 2010) across all recording pairs. We found no correlations between both measured across the whole population of matched spike-spike and LFP-LFP pairs ($p=0.444$, $r=-0.019$, $n=1597$), or when restricting the analysis to those LFP-LFP pairs that showed significant theta-gamma correlation at the single pair level ($p=0.506$, $r=-0.059$ $n=128$). In a second analysis, we correlated the normalized spiketrain correlations with the change in theta-gamma LFP correlation from pre- to post-attention cue. Similar to the previous analysis, spiketrain correlation and LFP cross frequency pairs were not correlated across the whole population ($p=0.554$, $r=-0.015$, $n=1597$), and neither for the subset of LFP pairs with a significantly increased theta-gamma correlation ($p=0.860$, $r=0.016$, $n=128$). These results quantify that spiketrain correlations of neurons and LFP theta-gamma phase amplitude correlations are largely independent phenomena while inter-relationships could be explored in more detail in future studies.

2.5 Discussion

We found that neurons in anterior cingulate and prefrontal cortices show functionally correlated spiking activity during covert stimulus selection. Using smoothed JPSTH analysis we showed that task-modulated cell pairs showed correlated firing when spiketrains were smoothed by 50-200ms. Dynamic crosscorrelations in this time window showed five major characteristics: Firstly, they emerged independently of the absolute firing rate of neurons. Secondly, spiketrain correlations were evident across cell classes. Putative pyramidal cells showed positive correlations with multiunit activity, as well as with putative interneurons across area boundaries. Thirdly, spiketrain correlations were anatomically specific, becoming evident for only a subset of area combinations. The strongest correlations were found between cells in ACC and dorsal PFC (dPFC, area 8 and 9). Fourth, spiketrain correlations carried attention

information as evident in cell pairs engaging in correlations when both cells preferred contra- over ipsilateral attention shifts. Lastly, we found that firing events in dPFC tended to precede firing events in ACC as shown in an average asymmetry of coincident firing. Taken together, these findings provide direct evidence for spiketrain correlations between two main top-down control structures that bias and implement the selection of relevant target stimuli during goal-directed behavior.

Connectivity in fronto-cingulate cortex

The prefrontal and cingulate cortices are interconnected anatomically (Medalla and Barbas, 2009, 2010), as well as functionally (Tsujimoto et al., 2006, 2010; Cavanagh et al., 2009; Rothé et al., 2011; Voloh et al., 2015). However, most reports are not directly linked to attention shifts. Recently, we showed that, specifically in the time window of active deployment of attention, burst spiking in ACC phase-locked to LFPs in PFC at beta and gamma band frequencies (Womelsdorf et al., 2014a). These data suggest substantial functional interactions between ACC and PFC; with our study, we add to this by showing that single neurons correlate their firing in ACC and PFC at a relatively slow 50-200ms time scale during covert shifts of attention. This potentially identifies the neural basis of integrating relevant attention information in the fronto-cingulate network.

In our study, dPFC emerged as a functional hub for cross-area communication during the attentional shift period. Neurons between ACC and dPFC as well as between vmPFC and dPFC synchronized their firing when attention was shifted. In addition, neurons within dPFC strongly synchronized during this time (Figure 2.3D). All these areas have previously been implicated in unique functional contributions during attentional control (Rowe and Passingham, 2001; Lebedev et al., 2004; Hayden et al., 2011; Kaping et al., 2011; Shenhav et al., 2013). An increased communication between ACC and dPFC may reflect the integration of spatial information of the current target as well as the distracter. This is supported by our finding that many ACC-dPFC cell pairs show spatial selectivity in their firing (Figure 2.5B). DPFC also showed spiketrain correlations with vmPFC, which is likely relevant for the integration of value (here:

'relevance-') information and target location, potentially allowing a successful shift of attention to the target independent of its value (Kaping et al., 2011; Rudorf and Hare, 2014). Generally, spatial information was prominently encoded in these cell-cell correlations, since neurons that encoded a contralateral attention shift were significantly more likely to engage in functional synchronization (Figure 2.5A). Importantly, these cell-cell communications are specifically associated with the process of covertly shifting attention, as increases in correlated firing were not apparent after the color cue onset (Figure 2.3D).

Time course of neuronal correlations in PFC/ACC

We observed the increase in spiketrain correlations between cells across the PFC/ACC only when JPSTHs were smoothed with 50-200ms windows, but not with smaller or larger windows (Figure 2.1C, D, Figure 2.3C). Functional correlations on faster time scales would suggest direct monosynaptic connections. However, the probability of those decreased significantly once neurons were separated by more than 200µm (e.g. Fujisawa et al., 2008). The larger timescale that is necessary for our effect to be observed suggests that neuron pairs were not monosynaptically connected.

Furthermore, firing correlations between cells emerged approximately 250-300ms after the animal was cued to shift attention. This time course of interaction fits well with our and other reports showing that spatial attention information in PFC/ACC emerges around 300ms after the appearance of the attention cue across the whole medial to lateral extent of the PFC (Kaping et al., 2011; Rainer et al., 1998; Suzuki and Gottlieb, 2013). This temporal coincidence suggests that attention information in the firing of ACC/PFC neurons translates into correlations for subsets of cells in ACC and dPFC.

Asymmetry of firing correlations may indicate information flow from dPFC to ACC

The asymmetry index analysis revealed that firing events in dPFC significantly preceded firing events in ACC after attention cue onset (Figure 2.4B, C). This suggests that when attention is covertly shifted,

information that is shared or transferred between ACC and dPFC appears earlier in dPFC. This could reflect the transfer of cue-to-target mapping rule information from dPFC to ACC. The attention shift is cued by the fixation point color, corresponding to two different rules. Previous studies have shown that rule-specific information is encoded in PFC (Wallis et al., 2001; Bongard and Nieder, 2010; Buschman et al., 2012), and also ACC where it lags behind PFC for over-learned tasks, and precedes PFC during rapid adaption of task rules (Johnston et al., 2007; Womelsdorf et al., 2010). The precedence of spiking in dPFC over ACC in our study is consistent with these earlier findings. We thus speculate that transfer of rule information from PFC to ACC could enhance target selection and sustaining attentional focus in light of distracters (Kaping et al., 2011; Shenhav et al., 2013).

Inter-areal spiketrain correlations are subserved by putative pyramidal cells and interneurons

We focused our analysis on functional correlations of neuronal spiketrains at slow time scales. In this window, spiketrain correlations may indicate polysynaptic (possibly dendritic) interactions of single cells with the local population of cells that can be indexed as multiunit response. Consistent with this prediction we could trace back significant spiketrain correlation to mixed multiunit signals with a population of broad-spiking, putative pyramidal cells. Notably, this single cell to multiunit correlation was robust for inter-areal comparisons, providing quantitative evidence that the correlation of cells with net excitatory influence supports attention shifts (Figure 2.6E). Horizontal connections between distant pyramidal cells may be the source of this net correlation (Pucak and Levitt, 1996; González-Burgos et al., 2000). Potentially supporting recurrent activation, such connections have been proposed to play an important role during working memory delay activity (González-Burgos et al., 2000) and are particularly extensive in primate PFC (Elston, 2003; Elston et al., 2005). Together, these findings provide further insights into functional recruitments of cell classes during selective attentional processes (Ardid et al., 2015).

Processes underlying spiketrain correlations

The correlations observed between cells across the PFC/ACC can conceptually result from different scenarios (Figure 2.7). Correlated firing could be caused by recurrent activity in the network, whereby one area influences another, allowing communication of relevant attention information (Figure 2.7A). Alternatively, spiketrain correlations could be caused by a common third-party influence on two subareas in the PFC/ACC concurrently (Figure 2.7B). These two scenarios are difficult to distinguish. In addition, several factors can contribute to increased dynamic crosscorrelations. Excitability covariations, such as independent increases in the excitability of both cells, as well as latency covariations, which are trial-to-trial covariations of spike timing, can also contribute to a peak in the JPSTH (Brody, 1999a, 1999b). Additionally, the covariation of stereotyped neuronal transients to external events can also induce a peak in the JPSTH (Friston, 1995; but also see Vaadia et al., 1995a). Considering the slow time scale of the correlations we observe and the distance between neurons, monosynaptic connections can be ruled out as a major source (Fujisawa et al., 2008). Further supporting this reasoning, analysis of crosscorrelations on a fast time scale (5ms resolution) showed no apparent correlation linked to attention shifts (data not shown). In line with this, the observed spiketrain correlations between ACC/PFC neurons may best be understood as excitability covariations of distant neurons. These could be caused by either direct recurrent activity (Figure 2.7A) or by common input from a third area (Figure 2.7B). Third-party input to both ACC and PFC could be realized by several candidate mechanisms, including dopaminergic, cholinergic and noradrenergic neuromodulation (Lapiz and Morilak, 2006; Fujisawa and Buzsáki, 2011; Hasselmo and Sarter, 2011; Sara and Bouret, 2012).

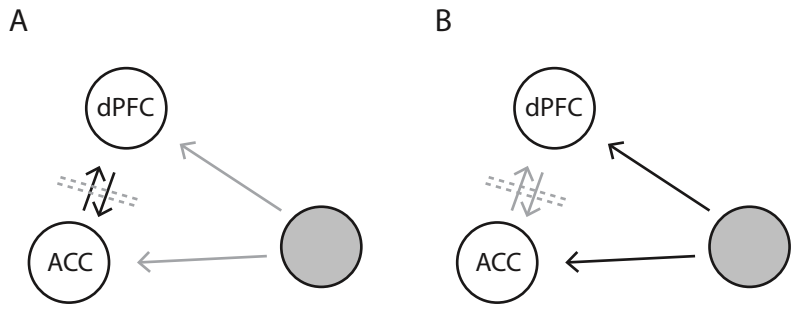

Figure 2.7 Sources of correlated firing of cells pairs from different cortical fields in ACC and dorsal PFC.

(A) Correlated firing could be caused by recurrent activity in the network that communicates relevant attention information. (**B**) Correlated firing could be caused by a third-party area that concurrently influences ACC/PFC.

Conclusions

We found direct evidence that correlated activity between neurons across subfields of the fronto-cingulate cortex can carry relevant attention information. This effect is evident for neurons that encode similar spatial attention information in their firing rate modulations; whereby neurons encoding contralateral shifts engage in increased synchronization, while neurons encoding ipsilateral shifts engage in decreased synchronization. These correlations occurred on a slow, characteristic 50-200ms time scale, suggesting that the transient single cell coupling is not driven by monosynaptic connections. Consistent with this, we traced back spiketrain correlations to the population of broad-spiking, putative pyramidal cell. Together these findings reveal that functional correlations of firing between neurons in ACC and PFC support the interchange of goal-relevant information during attentional stimulus selection.

2.6 References

Aertsen AM, Gerstein GL, Habib MK, Palm G (1989) Dynamics of neuronal firing correlation: modulation of " effective connectivity ." J Neurophysiol 61:900–917.

Alexander WH, Brown JW (2011) Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14:1338–1344.

Ardid S, Vinck M, Kaping D, Marquez S, Everling S, Womelsdorf T (2015) Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. J Neurosci 35(7): 2975-2991.

Barbas H, Zikopoulos B (2007) The prefrontal cortex and flexible behavior. Neuroscientist 13:532–545.

Bongard S, Nieder A (2010) Basic mathematical rules are encoded by primate prefrontal cortex neurons. Proc Natl Acad Sci U S A 107:2277–2282.

Brody C (1999a) Correlations without synchrony. Neural Comput 11:1537–1551.

Brody C (1999b) Disambiguating different covariation types. Neural Comput 11:1527–1535.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PGF, Kwok SC, Phillips A, Tanaka K (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science 325:52–58.

Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK (2012) Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron 76:838–846.

Cavanagh JF, Cohen MX, Allen JJB (2009) Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. J Neurosci 29:98–105.

Cohen JY, Crowder E a, Heitz RP, Subraveti CR, Thompson KG, Woodman GF, Schall JD (2010) Cooperation and competition among frontal eye field neurons during visual target selection. J Neurosci 30:3227–3238.

Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. J Neurophysiol 88:3487–3497.

De la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448:802–806.

Diester I, Nieder A (2008) Complementary contributions of prefrontal neuron classes in abstract numerical categorization. J Neurosci 28:7737–7747.

Elston GN (2003) The pyramidal neuron in occipital, temporal and prefrontal cortex of the owl monkey (Aotus trivirgatus): Regional specialization in cell structure. Eur J Neurosci 17:1313–1318.

Elston GN, Benavides-Piccione R, DeFelipe J (2005) A study of pyramidal cell structure in the cingulate cortex of the macaque monkey with comparative notes on inferotemporal and primary visual cortex. Cereb Cortex 15:64–73.

Friston K (1995) Neuronal transients. Proc R Soc London 261:401–405.

Fujisawa S, Amarasingham A, Harrison MT, Buzsáki G (2008) Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nat Neurosci 11:823–833.

Fujisawa S, Buzsáki G (2011) A 4 Hz oscillation adaptively synchronizes prefrontal, VTA, and hippocampal activities. Neuron 72:153–165.

Genovesio A, Wise SP, Passingham RE (2014) Prefrontal – parietal function: from foraging to foresight. Trends Cogn Sci 18:72–81.

González-Burgos G, Barrionuevo G, Lewis DA (2000) Horizontal synaptic connections in monkey prefrontal cortex: an in vitro electrophysiological study. Cereb Cortex 10:82–92.

Hasselmo ME, Sarter M (2011) Modes and models of forebrain cholinergic neuromodulation of cognition. Neuropsychopharmacology 36:52–73.

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci 31:4178–4187.

Johnston K, DeSouza JFX, Everling S (2009) Monkey prefrontal cortical pyramidal and putative interneurons exhibit differential patterns of activity between prosaccade and antisaccade tasks. J Neurosci 29:5516–5524.

Johnston K, Levin HM, Koval MJ, Everling S (2007) Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron 53:453–462.

Joshua M, Adler A, Prut Y, Vaadia E, Wickens JR, Bergman H (2009) Synchronization of midbrain dopaminergic neurons is enhanced by rewarding events. Neuron 62:695–704.

Kaping D, Vinck M, Hutchison RM, Everling S, Womelsdorf T (2011) Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. PLoS Biol 9:e1001224.

Kennerley SW, Behrens TEJ, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581–1589.

Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. J Neurosci 25:3661–3673.

Lapiz MDS, Morilak DA (2006) Noradrenergic modulation of cognitive function in rat medial prefrontal cortex as measured by attentional set shifting capability. Neuroscience 137:1039–1049.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Medalla M, Barbas H (2009) Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. Neuron 61:609–620.

Medalla M, Barbas H (2010) Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. J Neurosci 30:16068–16081.

Miller EK, Buschman TJ (2012) Cortical circuits for the control of attention. Curr Opin Neurobiol:1–7.

Passingham R, Wise SP (2012) The Neurobiology of the Prefrontal Cotex: Anatomy, Evolution, and the Origin of Insight. Oxford University Press.

Paz R, Bauer EP, Paré D (2007) Learning-related facilitation of rhinal interactions by medial prefrontal inputs. J Neurosci 27:6542–6551.

Petrides M, Pandya DN (1994) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and macaque frontal cortex. In: Handbook of neuropsychology Vol IX, pp 17–58. Amsterdam: Elsevier.

Petrides M, Pandya DN (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. Eur J Neurosci 11:1011–1036.

Phillips JM, Vinck M, Everling S, Womelsdorf T (2013) A Long-Range Fronto-Parietal 5- to 10-Hz Network Predicts “Top-Down” Controlled Guidance in a Task-Switch Paradigm. Cereb Cortex 24:1996–2008.

Pucak M, Levitt J (1996) Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. J Comp Neurol 376:614–630.

Rainer G, Asaad WF, Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature 393:577–579.

Reid R, Alonso J-M (1995) Specificity of monosynaptic connections from thalamus to visual cortex. Nature 378:281–284.

Rosset A, Spadola L, Ratib O (2004) OsiriX: an open-source software for navigating in multidimensional DICOM images. J Digit Imaging 17:205–216.

Rothé M, Quilodran R, Sallet J, Procyk E (2011) Coordination of high gamma activity in anterior cingulate and lateral prefrontal cortical areas during adaptation. J Neurosci 31:11110–11117.

Rowe JB, Passingham RE (2001) Working memory for location and time: activity in prefrontal area 46 relates to selection rather than maintenance in memory. Neuroimage 14:77–86.

Rudorf S, Hare T (2014) Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice. J Neurosci 34:15988–15996.

Saleem KS, Kondo H, Price JL (2008) Complementary Circuits Connecting the Orbital and Medial Prefrontal Networks with the Temporal, Insular, and Opercular Cortex in the Macaque Monkey. J Comp Neurol 506:659–693.

Saleem KS, Miller B, Price JL (2014) Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. J Comp Neurol 522:1641–1690.

Sallet J, Mars RB, Noonan MP, Neubert F-X, Jbabdi S, O'Reilly JX, Filippini N, Thomas AG, Rushworth MF (2013) The organization of dorsal frontal cortex in humans and macaques. J Neurosci 33:12255–12274.

Sara SJ, Bouret S (2012) Orienting and reorienting: the locus coeruleus mediates cognition through arousal. Neuron 76:130–141.

Schultz SR, Kitamura K, Post-Uiterweer A, Krupic J, Häusser M (2009) Spatial pattern coding of sensory information by climbing fiber-evoked calcium signals in networks of neighboring cerebellar Purkinje cells. J Neurosci 29:8005–8015.

Shen C, Ardid S, Kaping D, Westendorff S, Everling S, Womelsdorf T (2014) Anterior Cingulate Cortex Cells Identify Process-Specific Errors of Attentional Control Prior to Transient Prefrontal-Cingulate Inhibition. Cereb Cortex.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Suzuki M, Gottlieb J (2013) Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. Nat Neurosci 16:98–U139.

Tort AB, Komorowski R, Eichenbaum H, & Kopell N (2010) Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. Journal of neurophysiol-ogy 104(2):1195-1210.

Tsujimoto T, Shimazu H, Isomura Y (2006) Direct recording of theta oscillations in primate prefrontal and anterior cingulate cortices. J Neurophysiol 95:2987–3000.

Tsujimoto T, Shimazu H, Isomura Y, Sasaki K (2010) Theta oscillations in primate prefrontal and anterior cingulate cortices in forewarned reaction time tasks. J Neurophysiol 103:827–843.

Vaadia E, Aertsen AM, Nelken I (1995a) "Dynamics of neuronal interactions" cannot be explained by "neuronal transients." Proc R Soc London 261:407–410.

Vaadia E, Haalman I, Abeles M, Bergman H, Prut Y, Slovin H, Aertsen AM (1995b) Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. Nature 373:515–518.

Van Essen D (2002) Windows on the brain: the emerging role of atlases and databases in neuroscience. Curr Opin Neurobiol 12:574–579.

Voloh B, Valiante TA, Everling S, Womelsdorf T (2015) Theta-gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. Proc Natl Acad Sci U S A 112:8457-8462.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. Nature 411:953–956.

Womelsdorf T, Ardid S, Everling S, Valiante TA (2014a) Burst Firing Synchronizes Prefrontal and Anterior Cingulate Cortex during Attentional Control. Curr Biol:1–9.

Womelsdorf T, Valiante TA, Sahin NT, Miller KJ, Tiesinga P (2014b) Dynamic circuit motifs underlying rhythmic gain control, gating and integration. Nat Neurosci 17:1031–1039.

Womelsdorf T, Vinck M, Leung LS, Everling S (2010) Selective theta-synchronization of choice-relevant information subserves goal-directed behavior. Front Hum Neurosci 4:210.

Zhou X, Zhu D, Katsuki F, Qi XL, Lees CJ, Bennett AJ, Salinas E, Stanford TR, C Constantinidis C (2014) Age-dependent changes in prefrontal intrinsic connectivity. Proc. Natl. Acad. Sci. USA 111(10):3853-3858.

Zhuang KZ, Lebedev MA, Nicolelis MA. Joint cross-correlation analysis reveals complex, time-dependent functional relationship between cortical neurons and arm electromyograms. J Neurophysiol. 2014 Dec 1;112(11):2865-87.

Chapter 3

Signed and unsigned prediction error signals in macaque fronto-striatal circuit carry information about the preceding feature choice

Mariann Oemisch[1], Stephanie Westendorff[1,2], Samira Azimi[1], Ali Hassani[1], Salva Ardid[3], Paul Tiesinga[4], Thilo Womelsdorf[1,5]

[1]Department of Biology, Centre for Vision Research, York University, Toronto, Ontario M6J 1P3, Canada; [2]Department of Animal Physiology, Institute of Neurobiology, University of Tübingen, Tübingen 72076, Germany; [3]Department of Mathematics, Boston University, Boston, MA 02215, USA; [4]Department of Neuroinformatics, Donders Centre for Neuroscience, Radboud University Nijmegen, Nijmegen, AJ 6525, The Netherlands; [5]Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240, USA

3.1 Summary

For classical reinforcement learning to take place, an error term is required. This error term reflects the discrepancy between the expected and the received outcome and is used to update value predictions, so that future choices are more likely to lead to positive rewarding outcomes. When only one stimulus feature among many predicts reward, it is pivotal for an agent to learn which specific feature is causally linked to changes in reward outcomes. This link could be achieved with feature-specific prediction error signals, but it has been unclear so far whether prediction errors (PEs) can carry sufficiently detailed information to serve as selective learning signals. We used a reinforcement learning model to identify trial-by-trial PEs in a feature-based reversal learning task performed by two macaques. We correlated neural activity in anterior cingulate (ACC) and prefrontal cortex (PFC), as well as in caudate nucleus and ventral striatum (VS) with trial-by-trial PEs based on the specific features that were selected in the preceding choice. We found that i) the majority of PE encoding neurons carried feature information from the preceding choice, and these signals were encoded more strongly than non-specific PEs, ii) feature-specific unsigned PEs that were uninformative about the valence of the outcome were most prevalent in all areas tested and were encoded for all feature types independent of reward-relevance, iii) feature-specific positive and negative PEs, in contrast, were valence-informative and selectively encoded for the reward-relevant dimension (color), whereby color-specific positive PEs were found in all areas and color-specific negative PEs consistently in VS, iv) PFC was the only area that encoded positive PEs consistently for all stimulus features, and v) the prevalent encoding of feature-specific PEs co-occurred with a propensity to multiplex task-relevant variables with performance information. These results document that across the fronto-striatal circuit, PE encoding is tuned to the specific features that give rise to the PE. We speculate that the feature specificity of the error signals allows to identify those synapses across a network that need updating during learning, enabling selective credit assignment in feature space.

3.2 Introduction

A hallmark of behavioral control seems to be to make accurate predictions about one's environment with the goal of maximizing future returns and minimizing future losses. Because we live in a changing and stochastic environment, predictions must continuously be monitored and are updated when expectations for those predictions are unmet (Dayan and Niv, 2008). The difference between a predicted outcome and one experienced is referred to as a prediction error (PE). This PE is then used to update the value prediction of those actions or stimuli that had been performed or selected, respectively, and that presumably led to the unexpected outcome (Sutton and Barto, 1998). This computation seems relatively straight forward when, for instance, a single well-defined action was performed, however, in the case that only one stimulus feature among many predicts reward, it is pivotal for an agent to learn which specific feature is causally linked to changes in reward outcome. This concept relates to the credit assignment problem, which describes the computational problem of correctly assigning credit of an outcome to the source of that outcome in the preceding choice. The encoding of credit assignment has been suggested to take place in the orbital frontal cortex (OFC) (Chau et al., 2015; Jocham et al., 2016; Noonan et al., 2017), and recently also in medial prefrontal cortex (PFC) and dorsolateral PFC (Akaishi et al., 2016; Asaad et al., 2017). The purpose of this study was to probe whether PE signals in several fronto-striatal regions could carry sufficiently detailed information about the specific features that had been chosen, to serve as selective learning signals. The computation of predictions and prediction errors is wide spread across the brain and observed in many modalities ranging from sensory perception (Akatsuka et al., 2007; Summerfield and Koechlin, 2008), to action (Matsumoto et al., 2007; Bestmann et al., 2008), language (Kutas and Hillyard, 1980), cognitive control (Seo and Lee, 2007; Alexander and Brown, 2011) and value processing (Schultz, 1998; Daw et al., 2011; Bach et al., 2017). In part, due to the nature of their different modalities, prediction errors can be distinguished into unsigned PEs, that reflect a term of surprise about a given outcome or event, and that are uninformative about the valence of the respective outcome. Other PEs, often observed in the domain of motivational value processing (Schultz, 1998), are signed and

therefore carry explicit information about the valence of the outcome. A neural correlate of these latter types of PEs has first been described in dopamine neurons of the midbrain (Schultz et al., 1993). These neurons signal a positive PE, by firing action potentials at the time of an unexpected reward delivery; a response that declines in magnitude when the reward becomes more predictable and ceases once the reward delivery is fully predicted (Fiorillo et al., 2003; Bayer and Glimcher, 2005). Likewise, dopamine neurons cease to fire action potentials when an expected reward delivery is omitted, thereby also signaling a negative PE (Nomoto et al., 2010). Across areas of the fronto-striatal system, which (not exclusively) includes PFC, anterior cingulate cortex (ACC), caudate nucleus (CD) and ventral striatum (VS), signed and unsigned PEs have since been observed linked to primary rewards (Hare et al., 2008; Asaad and Eskandar, 2011; Hayden et al., 2011; Kennerley et al., 2011), to action values (Matsumoto et al., 2007; Sul et al., 2010), to abstract decision states (Gläscher et al., 2010; Daw et al., 2011), to abstract rule violations (Corlett et al., 2004), and linked to a time average of PEs (Wittmann et al., 2016). Many of those regions that have been shown to encode PEs also encode information about relevant actions, stimuli and features. The VS and CD have been implicated in encoding the values of relevant stimuli and actions (Samejima et al., 2005; Lau and Glimcher, 2008; Kim and Hikosaka, 2013; Vo et al., 2014; Rothenhoefer et al., 2017). Neurons in PFC have been shown to encode information about task relevant stimuli, features, and locations (Everling et al., 2002; Lebedev et al., 2004; Donahue and Lee, 2015; Siegel et al., 2015), while neurons in ACC are implicated in encoding values of relevant stimulus features and actions (Rushworth et al., 2004; Kaping et al., 2011), suggesting these regions could potentially multiplex information about unexpected outcomes and feature choices. So far, it has remained unclear whether PE encoding can be specific with regards to the feature choice that preceded it. A recent study suggests that this may be the case in dorsolateral PFC neurons, which signaled information about the surprise and valence of the outcome, based on the previous trial's outcome, as well as information about the stimulus/location that was chosen (Asaad et al., 2017). However, what has remained unknown is i) whether PE signals can be selective for one of several features of a given stimulus, and if so ii) whether

these signals carry information about valence and relevance of the specific feature, iii) whether these signals are selectively found in the PFC or also in other regions of the fronto-striatal circuit that have been implicated in encoding PE signals, and iv) in which of these areas these signals emerge first.

3.3 Methods

3.3.1 Electrophysiological recordings

Data was collected from two male rhesus macaques (*Macaca mulatta*). All animal care and experimental protocols were approved by the York University Council on Animal Care and were in accordance with the Canadian Council on Animal Care guidelines. Extra-cellular recordings were made with 1-12 tungsten electrodes (impedance 1.2 - 2.2 MOhm, FHC, Bowdoinham, ME) in anterior cingulate cortex (area 24, ACC), prefrontal cortex (area 46, PFC), caudate nucleus (CD) and ventral striatum (VS) (Figure 3.1C) through rectangular recording chambers (20 by 25 mm) implanted over the right hemisphere. Recording locations were identified using MRI images obtained following initial chamber placement, and daily with audible profiles of spiking activity. Electrodes were lowered daily through guide tubes using software controlled precision micro-drives (NAN Instruments Ltd., Israel). Data amplification, filtering, and acquisition were done with a multi-channel acquisition processor (*Neuralynx*). Spiking activity was obtained following a 300 - 8,000 Hz passband filter and further amplification and digitization at 40 kHz sampling rate. Sorting and isolation of single unit activity was performed offline with Plexon Offline Sorter, based on principal component analysis of the spike waveforms. Experiments were performed in a custom-made sound attenuating isolation chamber. Monkeys sat in a custom-made primate chair viewing visual stimuli on a computer monitor (60Hz refresh rate, distance of 58cm). Eye positions were monitored using a video-based eye-tracking system (*EyeLink*) calibrated prior to each experiment to a 9-point fixation pattern. Eye fixation was controlled within a 1.4-2.0 degree radius window. During the experiments, stimulus presentation, monitored eye positions, and reward delivery were controlled via

MonkeyLogic (open-source software http://www.brown.edu/Research/monkeylogic/). Liquid reward was delivered by a custom-made, air-compression controlled, mechanical valve system.

3.3.2 Behavioral paradigm

The monkeys performed a feature-based reversal learning task that required covert spatial attention to one of two stimuli dependent on color-reward associations. These color-reward associations were reversed in an un-cued manner between blocks of trials with constant color-reward association (Figure 3.1A). By separating the location of attention from the location of the saccadic response, this task allowed an identification of neural responses to the location of attentional focus independent of neural signals linked to response preparations, during reversal learning. Each trial started with the appearance of a grey central fixation point, which the monkey had to fixate. After 0.5 - 0.9s, two black/white drifting gratings appeared to the left and right of the central fixation point. Following another 0.4s the two stimulus gratings either changed color to black/green and black/red (*monkey K*: black/cyan and black/yellow), or started moving in opposite directions up and down, followed after 0.5 - 0.9s by the onset of the second stimulus feature that had not been presented so far, e.g. if after 0.4s the grating stimuli changed color then after another 0.5 - 0.9s they started moving in opposite directions. After 0.4 - 1s either the red and green stimulus dimmed simultaneously for 0.3s or they dimmed separated by 0.55s, whereby either the red or green stimulus could dim first. The dimming represented the go-cue to make a saccade to one of two response targets displayed above and below the central fixation point (Figure 3.1A). Please note that the monkeys needed to keep central fixation until this dimming event occurred. A saccadic response following the dimming was only rewarded if it was made to the response target that corresponded to the movement direction of the stimulus with the color that was associated with reward in the current block of trials, e.g. if the red stimulus was the currently rewarded target and was moving upward, a saccade had to be made to the upper response target at the time the red stimulus dimmed. A saccadic response was not rewarded if it was made to the response target that corresponded to the movement direction of the

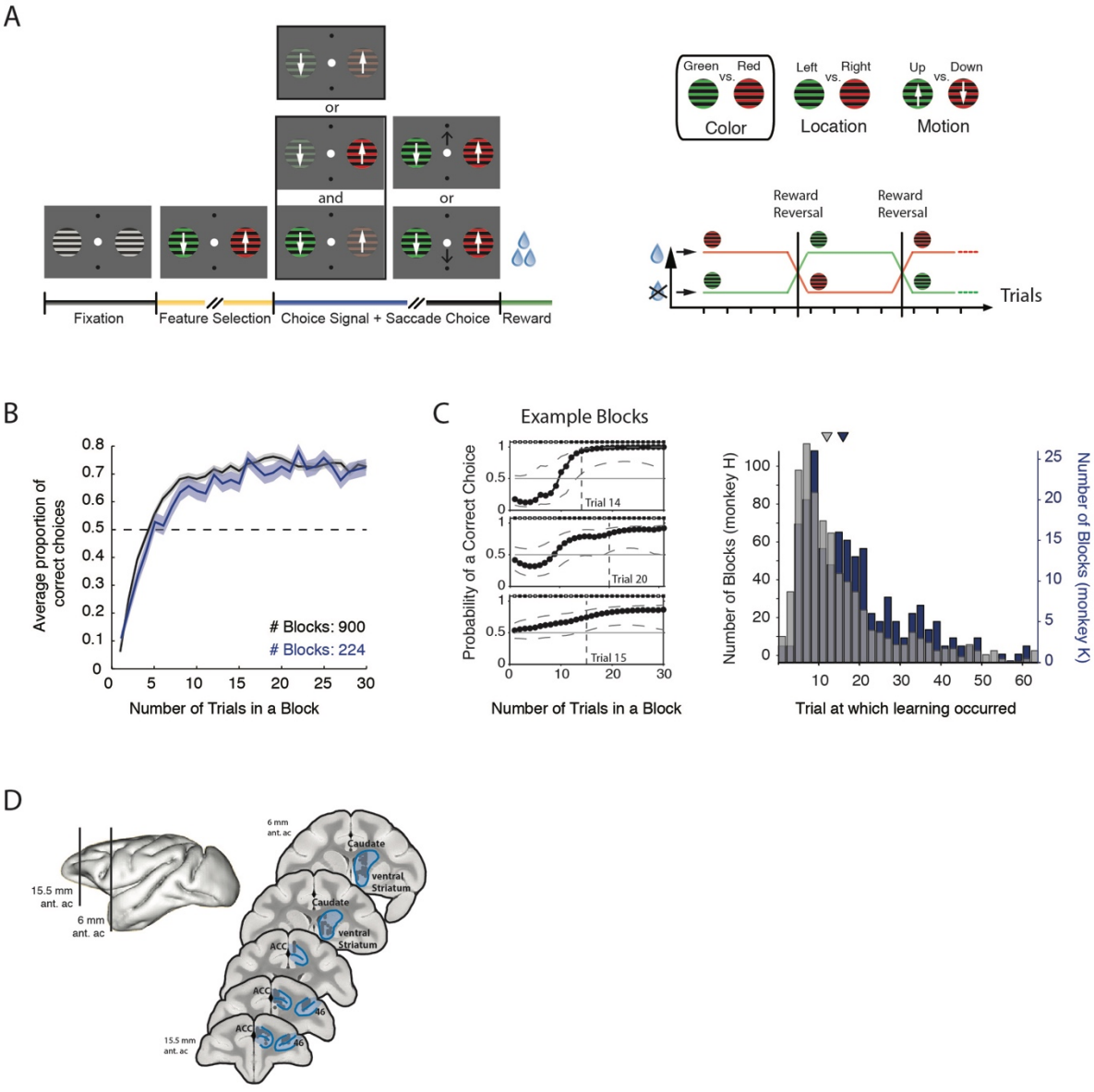

Figure 3.1 Behavioral task, performance and recording locations.

(A) Feature-based reversal learning task. *Left:* Animals are presented with two black/white stimulus gratings to the left and right of a central fixation point. The stimulus gratings then become colored and start moving in opposite directions. Dimming of the stimuli served as Choice/Go signal. At the time of the dimming of the target stimulus the animals had to indicate the motion direction of the target stimulus by making a corresponding up- or downward saccade in order to receive a liquid reward. Dimming of the target stimulus occurred either before, after or at the same time as the dimming of the distractor stimulus. *Right:* Three features characterize each stimulus – color, location, and motion direction. Only the color feature is directly linked to reward outcome. The task is a deterministic reversal learning task, whereby only one color at a time is rewarded. This reward contingency switches repeatedly and unannounced in a block-design fashion. **(B)** Average proportion of correct choices relative to the reversal for *monkey H*

(grey) and *monkey K* (blue). **(C)** *Left:* Three example blocks showing the probability of a correct choice as computed with the EM algorithm (see text for details). Dotted vertical lines and label indicate the trial at which learning was identified. Striped grey lines represent 95% confidence interval. Grey squares above each plot indicate error trials, black squares indicate correct trials. *Right:* Histogram of learning trials identified using the EM algorithm across reversal blocks for *monkey H* (grey) and *monkey K* (blue). Triangles indicate medians (*monkey H*: 12, *monkey K*: 16). **(D)** Illustration of recording locations relative to stereotaxic zero. Example slices show recording locations for *monkey H* as grey circles (see text for details).

stimulus with the non-reward associated color. A correct response was followed by 0.33ml of water delivered to the monkey's mouth. Across trials of a block the color-reward association remained constant for 30 to a maximum of 100 trials. Performance of 90% rewarded trials (calculated as running average over the last 12 trials) automatically induced a block change. The block change was un-cued, requiring the subject to use the reward outcome they received to learn when the color-reward association was reversed in order to covertly select the stimulus with the rewarded color. In contrast to color, other stimulus features (motion direction or stimulus location) were only randomly related to reward outcome. Saccadic responses had to be initialized within 0.5 s after dimming onset to be considered a choice (rewarded or non-rewarded). All other saccadic responses, e.g. towards the peripheral stimuli, were considered non-choice errors.

3.3.3 Stimuli

We used blocksine gratings with rounded-off edges for the peripheral stimuli (Figure 3.1A), moving within a circular aperture at 0.8 °/s and a spatial frequency of 1.2 (cycles/°) and a radius of 2.0°. Gratings were presented at 5° eccentricity to the left and right of the fixation point.

3.3.4 Data analysis

Analysis was performed with custom MATLAB code (Mathworks, Natick, MA), utilizing functions from the open-source Fieldtrip toolbox (http://www.ru.nl/fcdonders/fieldtrip/). All spike-density functions were

smoothed with a Gaussian kernel with a standard deviation of 25ms. Only units with a minimum firing rate of 0.5Hz within the feedback epoch were analyzed. Only correct and incorrect choice trials were analyzed, whereby correct choice trials were rewarded trials, while incorrect choice trials were either made to the non-rewarded stimulus or in the incorrect response time window (first vs. second dimming). Fixation breaks, early responses, and no-response trials were not included in any analyses.

Expectation maximization algorithm

To identify at which trial during a block the monkey showed statistically reliable learning we analyzed the monkeys' trial-by-trial choice dynamics using the state–space framework introduced by Smith and Brown (2003), and implemented by Smith et al., (2004). This framework entails a state equation that describes the internal learning process as a hidden Markov or latent process and is updated with each trial. The learning state process estimates the probability of a correct (rewarded) choice in each trial and thus provides the learning curve of subjects (see Figure 3.1C left for example blocks). The algorithm estimates learning from the perspective of an ideal observer that takes into account all trial outcomes of subjects' choices in a block of trials to estimate the probability that the outcome in a single trial is correct or incorrect. This probability is then used to calculate the confidence range of observing a correct response. We defined the learning trial as the earliest trial in a block at which the lower confidence bound of the probability for a correct response exceeded the $p = 0.5$ chance level.

More specifically, the algorithm defines the learning state process as a random walk whereby each trial's probability of a correct response depends on the previous trials probability, or on the chance level in case there was no previous trial's probability i.e. at the beginning of blocks. According to this formulation, the subject's choices across trials follow a random strategy. The mean of the random process reflects the current probability for a correct response. The variance of the random process determines how fast the learning state process can change from trial to trial and thus, how rapidly learning can take place (see Smith et al, 2004). The Expectation-Maximization (EM) algorithm is used to estimate the mean and

variance of the random process by maximum likelihood estimation to derive the probability to observe a correct response in each trial as a function of the trial number (Smith and Brown, 2003). A forward filter estimates the variance and mean of the value of the Gaussian Random Variable from the first trial to the current trial. This forward process reflects a state estimate from the perspective of the subject performing the task. An additional smoothing algorithm takes the perspective of an ideal observer and estimates the current trials mean and variance of the state process using data from all trials. The estimates of both, the forward filter and the smoothing process are then used to calculate the probability density for the correct response probability at each trial. Please see Smith et al. (2004) equations 2.1 to 2.4 for details. The aforementioned procedure provides the learning curve, i.e. it provides for each trial the probability of a correct response given the sequence of correct and incorrect choices of the monkey. To identify the first trial in a block at which an ideal observer knows with $p \geq 0.95$ confidence that learning has taken place, we calculated the lower confidence bound and identified the first trial where it exceeded the $p = 0.5$ chance level, the first 'IO95' learning trial (see Smith et al. 2004). This corresponds to a 0.95 confidence level for an ideal observer to identify learning. The identification of a learning trial allowed the separation of trials into *during* learning and *after* learning trials.

Variable encoding in the outcome epoch

To characterize neural responses in the outcome epoch, we adapted analyses from Padoa-Schioppa and colleagues (Padoa-Schioppa and Assad, 2006, 2008; Cai and Padoa-Schioppa, 2014). We tested whether neurons encoded any of twelve variables at the time of reward onset/omission, which were selected based on our task demands and based on variables that have been shown to be encoded in prefrontal regions in previous literature. These twelve variables included the three stimulus features (color, location, motion) i) selected in the current choice independent of choice outcome (correct and error) (**chosen color, chosen location, chosen motion**) (Genovesio et al., 2014), ii) selected in the previous choice (trial n-1) independent of choice outcome (correct and error) (**previous chosen color, previous chosen location,**

previous chosen motion) (Genovesio et al., 2014; Donahue and Lee, 2015), iii) of the target independent of choice (correct and error) (**target color, target location, target motion**) (Westendorff et al., 2016), in addition to the variables **outcome** (correct and error), **previous outcome** (correct and error) (e.g. Donahue and Lee, 2015) and **learning progress** (correct trials *during* learning versus *after* learning as obtained from the EM algorithm described above).

To estimate the correlation between variables, we computed the correlation coefficient between any two trial vectors of the variables per recording session and then computed the average absolute correlation coefficient across recording sessions (the average correlation coefficient now varies between 0 and 1). The correlation matrix is shown in Figure S1A, Appendix A.

To identify whether any neuron encodes any one variable, we performed independent linear regressions for each neuron on each variable. Each regression was corrected by weighting the residual associated with each data point with a term proportional to 1/standard deviation of the distribution from which a data point is drawn. A neuron's firing rate was averaged in the 0.1 - 0.7 seconds after reward onset/omission and was considered to significantly encode a variable at $p \leq 0.05$. In general, a neuron's response could be explained by multiple variables, which is likely because variables are correlated with each other, a situation referred to as multi-collinearity. We therefore adapted two methods of variable selection used in the case of multi-linear regressions, the "best-subset" method and the "stepwise selection" method (Supplementary methods) (Dunn and Clark, 1987; Glantz and Slinker, 2001; Padoa-Schioppa and Assad, 2006).

Best-subset method. We computed for each subset of d variables the total number of neural responses explained by that subset and determined which subset explained the maximum number of responses. This was repeated for d=1, 2, 3.. variables per subset. We determined the number of variables necessary to characterize the population when 85% of the maximum number of responses explained was reached.

Both, the stepwise and the best-subset method assume that each neuron only encodes a single variable. We therefore tested for second-order encoding to determine the proportion of neurons that encoded more than one variable (Padoa-Schioppa and Assad, 2006).

Second order encoding. We found for each neuron the best-fit variable and its corresponding R^2 value. To determine whether adding an additional variable to the regression led to a significantly higher R^2 value, we computed:

$$F_{X,Y}=\frac{(n-3)*(R^2_{XY}-R^2_X)}{(1-R^2_{XY})} \quad \text{(eq. 1)}$$

where R^2_X is from the original linear regression on X only, R^2_{XY} is from the bilinear regression on X and Y and n is the number of trials. $F_{X,Y}$ is computed for each of the eleven possible second variables and the maximum F is found. If the corresponding p-value for the maximum F value is ≤ 0.01, we consider the neuron to significantly encode the second variable (see Supplementary Methods, Padoa-Schioppa and Assad, 2006). We also determined for each neuron whether encoding at the second order was linked to a significant interaction between the two variables at $p \leq 0.01$.

Prediction error signals

For the prediction error (PE) analysis, neurons were only included in the analysis if prediction errors could be computed for a minimum of 40 trials.

Quantifying prediction errors with reinforcement learning modeling. We quantified the trial-by-trial progression of PEs during reversal performance using a computational model that combines reinforcement learning (RL) principles with Bayesian tracking of reward probabilities for target features. This hybrid Bayesian-RL model was introduced before (Wilson and Niv, 2012; Niv et al., 2015) to account for behavioral adjustments of choices in a multidimensional visual learning task and was recently

validated as a model accounting for feature-based reversal learning in the macaque (Hassani et al., 2017). The model represents the stimuli in terms of their stimulus features (color, motion, location), feature values (color A, color B, downward motion, upward motion, left, right), and the actual combinations of feature values for stimulus 1 and stimulus 2.

The model uses Bayesian inference about which stimulus feature f (color, motion or location) is the likely target feature via $p(f|\mathcal{D}_{1:t})$ to obtain a feature-weighted representation for each stimulus. For tracking target feature probability, we denote the feature dimension as f_d (1: location, 2: direction of motion, 3: color) and for each d, f_d, takes two values 1 and 2. For instance, f_3=1 indicates the first color. We can then calculate the probability for the rewarded stimulus (the target) to have feature d, $p_d = p(d|\mathcal{D}_{1:t}) = \sum_{f_d=1,2} p(f_d|\mathcal{D}_{1:t})$. This defines a feature dimension weight $\phi_d = \frac{p_d^\alpha}{\sum_{d'} p_{d'}^\alpha}$, with exponent α and normalized to yield a sum across dimensions equal to one. The predicted reward value of a feature value is then denoted by W_{f_d} and scaled by the dimensional weight ϕ_d. The value of the specific stimulus i is given by the sum across all weighted feature values that are part of the stimulus

$$V_i = \sum_d \phi_d W_{f_d} \quad \text{(eq. 2)}$$

The choice of which stimulus is selected on a given trial is implemented with a softmax rule using the Boltzmann function

$$P(C_{t+1} = i) = \frac{\exp(\beta V_{i,t})}{\sum_j \exp(\beta V_{j,t})} \quad \text{(eq. 3)}$$

Following a choice, the stimulus values of the chosen stimulus are updated by a reward PE scaled by learning rate η according to:

$$W_{f_d,t+1} = W_{f_d,t} + \eta\left(R_t - V_{i,t}\right) \quad \text{(eq. 4)}$$

Values of the unchosen stimulus feature values were scaled down (decayed) by $(1-\omega)$, similar to previous studies (see Hassani et al., 2017; Niv et al., 2015), according to:

$$W_{f_d,t+1} = (1-\omega)W_{f_d,t} \quad \text{(eq. 5)}$$

In summary, feature values of the chosen stimulus are updated using the PE (eq. 4), and are separately scaled by a dimensional weight (that may be called attentional weight) calculated using Bayes updating of how the feature dimensions color, motion and location relate to reward outcomes (Figure 3.4A).

We optimized the model by minimizing the negative log likelihood over all trials using up to 20 iterations of the simplex optimization method (matlab function fminsearch) followed by fminunc which constructs derivative information. We used an 80% / 20% (training dataset / test dataset) cross-validation procedure repeated for n=100 times to quantify how well the model predicted the data. Each of the one hundred cross-validations optimized the model parameters on the training dataset. We then quantified the log-likelihood of the independent test dataset given the training datasets optimal parameter values. We validated that the described hybrid Bayes-RL model provides a better fit (lower log-likelihood and lowest Akaike Information Criterion) for the cross-validated test dataset than simpler models that either lacked the Bayesian dimension weighting, or that lacked the decay of nonchosen stimulus features (for a detailed evaluation of different models, see also Hassani et al., 2017).

Identification of prediction error encoding neurons. To identify PE encoding neurons, we correlated each neuron's firing rate time-resolved with PEs obtained from the RL model. Each correlation analysis required a minimum of 15 trials. We correlated firing rate with positive PEs in correct choice trials and with negative PEs in incorrect choice trials (see *Methods* for initial trial selection). To identify neurons that encoded an unsigned PE, we used partial correlation analysis to correlate firing rates with the absolute PE in correct and incorrect choice trials while partializing out the sign of the PE. The analysis

time ranged from -500 to 1500ms after the outcome event; time windows spanned 200ms and were shifted by 25ms. For a neuron to be considered to encode a non-specific positive, negative, or unsigned PE signal, it had to significantly positively correlate its firing rate with a positive, or unsigned PE, respectively (Spearman correlation, $p < 0.05$) (see Figure 3.4B for temporal evolution of PEs), for a minimum of four consecutive time bins following the outcome event, while not correlating positively in more than two consecutive time bins before the outcome event. For a neuron to be considered to encode a negative PE signal, it had to significantly negatively correlate its firing rate with a negative PE for a minimum of four consecutive time bins following the outcome event (Spearman correlation, $p < 0.05$), while not correlating negatively in more than two consecutive time bins before the outcome event.

To identify neurons that encoded a feature-specific PE signal, trials were split into the features of interest prior to the correlation analysis (color, location and motion direction, see Figure 3.1A). The principle for identifying positive, negative and unsigned feature-specific PE neurons was the same as for non-specific PE signals with additional criteria described in the following. For instance, for a neuron to be considered to encode a color-specific PE signal, it had to significantly encode a PE signal (as described above) in minimally four consecutive time bins for trials in which e.g. color 1 was chosen, while either not encoding or encoding significantly less a PE signal for trials in which color 2 was chosen. Significant differences between R values for the two trial types were computed by z-transforming R values and comparing them using a z-test:

$$Z_{observed} = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} \quad \text{(eq. 6)}$$

where z_1 and z_2 are the z-transformed R-values for the correlation with feature value 1 and feature value 2, respectively. When $Z_{observed}$ exceeded $|1.96|$ ($p<0.05$), R values were considered significantly different for a given time bin. In a minimum of four consecutive bins, R values from correlations with two different

feature values (e.g. color 1 chosen or color 2 chosen) had to significantly differ, while a PE had to be encoded for at least one of the two feature values according to the same criteria as for non-specific PE signals (see single neuron examples in Figure 3.6). The method of identification was the same for identifying location and motion-specific PE signals, with the exception of splitting trials according to chosen location or chosen motion direction, respectively. We determined for each neuron the duration in which it encoded a PE signal as the first span of four or more consecutive significant time bins after the feedback event.

Effect sizes of PE encoding (average R-values obtained across those time bins deemed as PE encoding for each cell) were compared using 2-way ANOVA with the factors PE type (positive, negative, unsigned) and area (ACC, PFC, CD, VS) for non-specific PEs, or using 3-way ANOVA with the factors PE type (positive, negative, unsigned), area (ACC, PFC, CD, VS) and PE feature (color, location, motion) for feature-specific PEs. Post-hoc analyses were done using t-tests ($\alpha = 0.05$). Proportions of PE signals observed per area were assessed using one-sided binomial tests, and comparisons of proportions were done using Chi-square tests with Bonferroni-Holm multiple comparison corrections.

The issue of multiple comparisons

Every neuron was correlated with PE signals a total of nine times (for three PE types and three feature types), which brings up the issue of multiple comparisons. The analyses of different PE types with regards to trial outcome were however independent, since correct (positive PE) and error (negative PE) trials did not overlap, and for the computation of unsigned PEs any effects of trial outcome were partialized out as a covariate. In addition, the three stimulus features (color, location, motion) were varied independently of each other and were not correlated (this was tested as part of the regression analysis described above, see Figure S1A, C Appendix A), which suggests that trial selections based on these variables can also be considered independent. We therefore refrained from applying multiple comparison corrections for the initial identification of PE encoding neurons. Proportions of PE signals observed per

area were tested against chance with one-sided binomial tests; for these analyses, every PE signal observed was only tested once, making the need for multiple comparisons unnecessary. Whenever we explicitly compared more than two measures or areas, we employed the Bonferroni-Holm multiple comparison correction.

Time courses of prediction error and trial outcome signals

To compare time courses of PE signals, as well as trial outcome signals, between areas, we determined for each neuron the first time window (minimum 4 consecutive bins) in which it encoded a PE / trial outcome signal significantly. We tested the resulting distributions for differences in the time courses (cumulative sums) with which the signals were encoded (Kolmogorov-Smirnoff test, Bonferroni-Holm multiple comparison correction, $\alpha = 0.05$). As an additional measure of latency for encoding PE's we tested whether the time point at which 25% of PE / trial outcome signals were encoded differed between areas using a randomization procedure ($\alpha = 0.05$).

Classification of cell-types according to action potential waveform characteristics

For the set of highly isolated neurons (*monkey H*: n = 428, *monkey K*: 111), we aligned, normalized, and averaged all action potentials (Ardid et al., 2015). To distinguish putative interneurons (narrow-spiking) and putative pyramidal cells (broad-spiking) in PFC and ACC, we analyzed the peak-to-trough duration and the time for repolarization for each neuron (Oemisch et al., 2015). The time for repolarization was defined as the time at which the waveform amplitude decayed 15% from its peak value. We computed the principal component analysis (PCA) and used the first component because it allowed for better discrimination between narrow- and broad-spiking cells, compared to any of the two measures alone (Hartigan dip test, $p < 0.0005$). In addition, a comparison of Akaike's and Bayesian was used to confirm that a two-Gaussian model fit the data better than a one-Gaussian model. To distinguish putative interneurons and putative medium-spiny neurons (MSNs) in CD and VS, we analyzed the peak width (at

half maximum) and *I*nitial *S*lope of *V*alley *D*ecay (ISVD, Berke, 2008; Lansink et al., 2010), as they provided a better waveform discrimination than e.g. peak-to-trough duration. The ISVD was computed as follows:

$$ISVD = 100 * \frac{(V_T - V_{0.26})}{A_{PT}} \quad \text{(eq. 7)}$$

where V_T is the most negative value (trough) of the spike waveform, $V_{0.26}$ is the voltage at 0.26 ms after V_T, and A_{PT} is the peak-to-trough amplitude (Lansink et al., 2010). Although we could not discard unimodality for the first PCA component (or for either of the single measures, Hartigan dip test, $p > 0.05$), Akaike's and Bayesian information criteria confirmed that a two-Gaussian model fit the data better than a one-Gaussian model. For frontal and cingulate units, we then used the two-Gaussian model and divided neurons into two groups of narrow and broad spiking units. For striatal units, because we could not discard unimodality for the first PCA component, we used the two-Gaussian model and defined two cutoffs that divided neurons into three groups. The first cutoff was defined as the point at which the likelihood of a narrow-spiking/putative interneuron was 3 times larger than the likelihood of a broad-spiking/putative principal cell, and vice versa for the second cutoff. We reliably classified PFC/ACC neurons (n = 385) as either putative pyramidal cells (broad spiking, n = 288, *monkey H*: 203, *monkey K*: 85) or putative interneurons (narrow-spiking, n = 97, *monkey H*: 78, *monkey K*: 19). Therefore, in *monkey H* 72% of neurons in ACC/PFC were identified as putative pyramidal cells while 28% of neurons were identified as putative interneurons. In *monkey K*, 82% of neurons were identified as putative pyramidal cells and 17% as putative interneurons. We classified 96% of striatal neurons (n = 221) as either putative MSNs (broad spiking, n = 158, *monkey H*: 96, *monkey K*: 62) or putative interneurons (narrow-spiking, n = 54, *monkey H*: 35, *monkey K*: 19), while n = 9 (*monkey H*: 8, *monkey K*: 1) neurons fell in between the criteria and could not be reliably classified. Therefore, in *monkey H* 73% of neurons in CD/VS were identified as putative MSNs while 27% of neurons were identified as putative interneurons. In *monkey K*, 77% of neurons were identified as putative MSNs and 23% as putative interneurons. For striatal units, we additionally verified our classification by comparing the firing rates between neurons classified as MSNs

and those classified as interneurons. Striatal interneurons tend to be fast-spiking interneurons and should have a higher firing rate than the relatively low-firing MSNs (Berke et al., 2004; Berke, 2008). Indeed, in both monkeys, neurons classified as interneurons had a higher mean firing rate (*monkey H*: 4.96±1.1 Hz, *monkey K*: 4.77 ± 2.62 Hz) than neurons classified as MSNs (*monkey H*: 1.70 ± 0.38 Hz, *monkey K*: 1.61 ± 0.26 Hz), and this was statistically reliable in both monkeys (t-test, *monkey H*: $p < 0.001$, *monkey K*: $p = 0.039$).

3.4 Results

We recorded a total of 563 units in the ACC (*monkey H*: 420, *monkey K*: 143), 439 units in PFC (*monkey H*: 320, *monkey K*: 119), 361 units in caudate nucleus (CD; *monkey H*: 233, *monkey K*: 128) and 239 units in ventral striatum (VS; *monkey H*: 166, *monkey K*: 73). 65% of neurons in *monkey H* and 70% of neurons in *monkey K* met the initial trial selection criteria for analysis (*see* Methods) and all following results are reported with regards to this neuron selection. Throughout recordings, both animals performed a feature-based reversal learning task (Figure 3.1A).

3.4.1 Behavioral performance

Average block performance for both animals is shown in Figure 3.1B (left). The average number of reversal blocks performed per recording session was 9.7±0.3 for *monkey H* and 8.5±0.5 for *monkey K*. The mean block length was 46±0.7 trials for *monkey H* (median = 37) and 46±1.3 trials for *monkey K* (median = 40). Using an expectation maximization (EM) algorithm to estimate learning and the trial at which learning occurred within a given reversal block (Figure 3.1C left) (see Methods, Smith et al., 2004), *monkey H* successfully learned an average of 78.7% of reversal blocks within a median of 12 trials (mean 16.2±0.4, Figure 3.1C right), and *monkey K* successfully learned an average of 88.5% of reversal blocks within a median of 16 trials (mean 19.5±0.8, Figure 3.1C right).

3.4.2 Task variables encoded in fronto-striatal areas

Overall, 78-79% of neurons encoded at least one of the task variables tested at the time of trial outcome. To identify which task variables were encoded by neurons in the outcome epoch, we adapted analyses from Padoa-Schioppa and colleagues (2006, 2008, 2014). For this analysis, neural responses across all areas in the outcome epoch (0.1 – 0.7 sec following outcome event onset) were regressed using single linear regressions against each variable tested (*see* Methods for details). In order to deal with collinearity between variables (Figure S1A, C Appendix A), and based on similar analyses by others (Padoa-Schioppa and Assad, 2006), we initially assumed for each neuron to only encode a single variable. Assigning each neuron only the variable that it encoded strongest (Regression analysis, largest R^2, $p \leq 0.05$), revealed that the largest fraction of neurons encoded trial outcome (*monkey H*: 33%, *monkey K*: 37%), with smaller fractions encoding trial outcome in the previous trial (*monkey H:* 8%, *monkey K:* 7%), learning progress (*monkey H:* 7%, *monkey K: 6%*), target color (*monkey H:* 7%, *monkey K: 7%*), chosen motion direction (*monkey H:* 7%, *monkey K: 4%*), chosen location (*monkey H: 3%, monkey K: 6%*), and so forth (see Figure 3.2A for all variables encoded in *monkey H* and *K*). The distribution of variables encoded was generally similar across areas (Figure S2 Appendix A), although we find that PFC encoded chosen motion more often than ACC or VS in *monkey H* and more often than CD and VS in *monkey K* (pairwise chi-square comparison, Bonferroni-Holm multiple comparison corrected, all $p < 0.05$), and chosen location was encoded more often in PFC compared with ACC in *monkey H* and compared with all other areas in *monkey K* (all $p < 0.05$). Additionally, in *monkey K* only, trial outcome was encoded more often in ACC compared with PFC ($p = 0.012$), and previous trial outcome was encoded more often in VS compared with PFC and CD (all $p < 0.05$).

To identify the selection of task variables that explained the largest number of neural responses, we adapted the best-subset method (see Supplementary results for stepwise-selection method). This revealed that the maximum percentage of responses that could be explained using all twelve variables was 41.9%

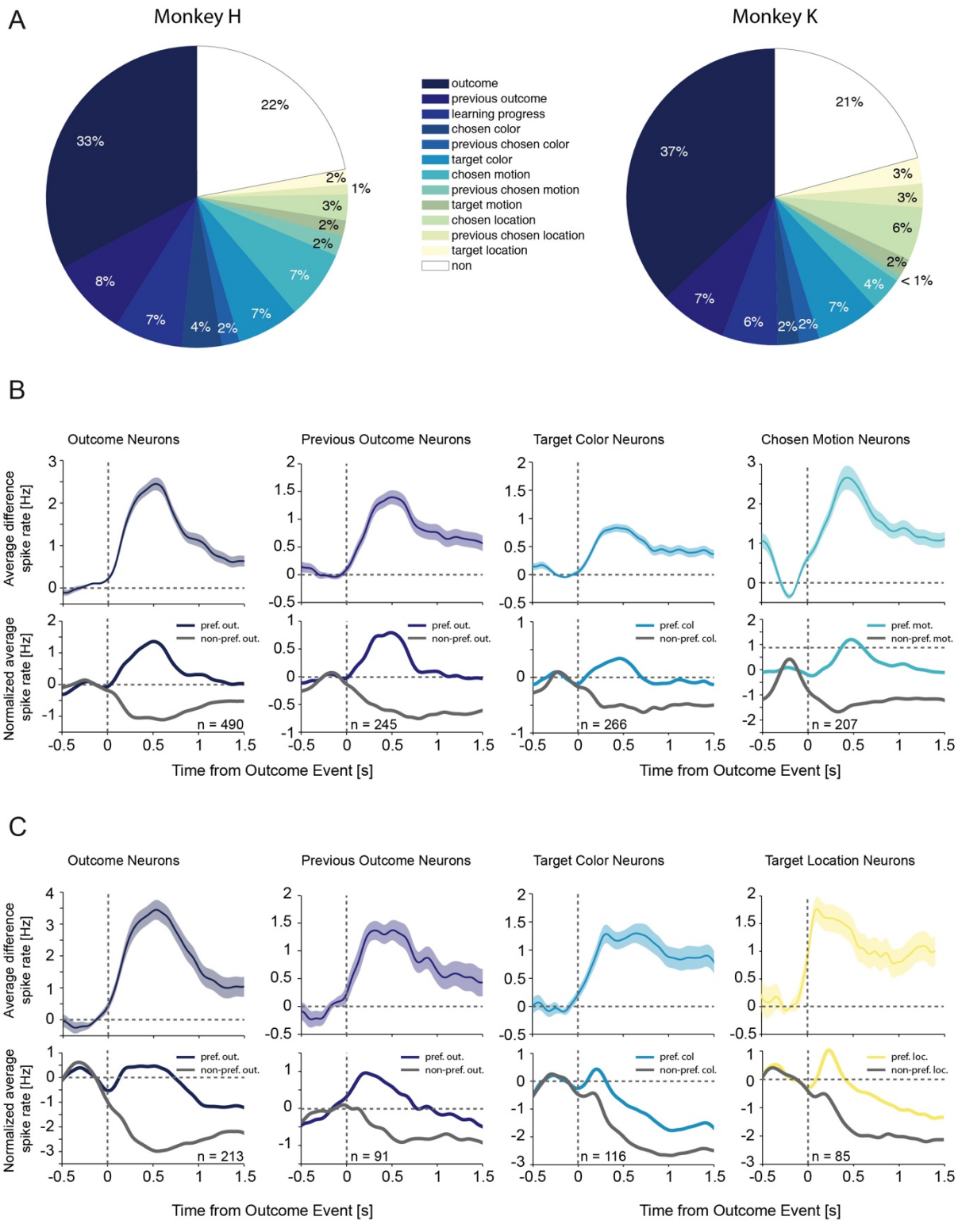

Figure 3.2 Task variables encoded at time of outcome.

(A) Proportion of task variables encoded at the time of outcome for *monkey H* (left) and *monkey K* (right) across all areas. Neurons are grouped based on the variable that provided the best fit for their responses (highest R, linear regression). **(B)** Average spike rate difference of neurons that encoded the four variables explaining the largest proportion of responses for *monkey H*. The bottom plot of each figure shows the average normalized spike rate for the two conditions for the given variable. Plotted are neurons that significantly encoded these variables. **(C)** Same as **B** for *monkey K*. Shaded error bars represent SEM. Number insets indicate numbers of neurons.

for *monkey H* and 40.9% for *monkey K*. 85% of this maximum number of responses could be explained with four variables (35.6% for *monkey H*, 34.7% for *monkey K*, Figure S1B, D Appendix A). This best subset of four variables contained the variables **outcome**, **previous outcome**, **target color, and chosen motion** for *monkey H*, and the variables **outcome**, **previous outcome**, **target color**, and **target location** for *monkey K*. Figure S1B and D Appendix A show the variables contained in the best subsets for subset sizes 1 to 5. The fact that the same sets of variables are included in best subsets of sizes 2, 3, 4 and 5 indicates the robustness of the results. Figure 3.2B and C show the mean difference firing rates across neurons encoding these four variables that combined explained the largest proportion of neural responses for *monkey H* and *K*, respectively.

The maximum percentage of responses explained was similar across areas ranging from 37 to 47% (ACC: *monkey H*: 45.5%, *monkey K*: 39.2%, PFC: *monkey H*: 37.0%, *monkey K*: 37.0%, CD: *monkey H*: 42.0%, *monkey K*: 47.1%, VS: *monkey H*: 46.0%, *monkey K*: 42.5%, pairwise chi-square comparisons with Bonferroni-Holm multiple comparison correction, all $p > 0.05$). For *monkey H*, 85% of the max. percentage of responses explained could be explained with five variables in ACC (*best subset*: outcome, previous outcome, target color, chosen motion, chosen color), four variables in PFC (*best subset*: outcome, previous outcome, target color, chosen motion), and three variables in caudate and ventral striatum each (CD *best subset*: outcome, target color, chosen motion; VS *best subset*: outcome, previous outcome, target color). For *monkey K*, 85% of the max. percentage of responses explained could be explained with three variables in ACC (*best subset*: outcome, previous outcome, target color), four

variables in PFC (*best subset*: outcome, target color, chosen location, target location), five variables in caudate (*best subset*: outcome, target color, learning progress, previous chosen color, target motion), and two variables in ventral striatum (*best subset*: outcome, previous outcome). This suggests that similar task variables explained the majority of responses in the areas tested here, although to somewhat differing degrees, and with *monkey H* tending to encode more motion-selective neurons and *monkey K* more location-selective neurons.

3.4.3 Multiplexing of task information

We observed that some cells encoded more than one variable. Therefore, after having examined single variable encoding in neurons in the previous section, we now examined multiplexing in single neurons. Some cells for instance encoded information about outcome and chosen color (Figure 3.3Aii) or target location (Figure 3.3Ai). Others encoded information about location of the target stimulus and previously chosen location (Figure 3.3Av). Figure 3.3C illustrates six of these example neurons from different brain areas that encoded different pairs of variables at the first and second order. To quantify this multiplexing we analyzed second order encoding using an approach described previously (Assad and Padoa-Schioppa, 2006). We found that 56% of neurons (*monkey H:* 53%, *monkey K*: 63%) that significantly encoded one task variables, also encoded a second task variable, which is more than expected by chance (binomial test, $p < .0001$). These were neurons for which a bi-linear regression with two predictor variables fit the data significantly better than a linear regression with only one predictor variable. Out of those neurons that significantly encoded a second variable 6.9% encoded an interaction between the two variables (*monkey H*: 7%, *monkey K*: 6.8%), suggesting that the majority of neurons multiplexed the two identified variables, rather than encoding an interaction between them. Second order encoding generally occurred in similar proportions across areas (ACC: 50% (*monkey H*) 62% (*monkey K*), PFC: 62% (*monkey H*) 67% (*monkey K*), CD: 55% (*monkey H*) 58% (*monkey K*), VS: 39% (*monkey H*) 70% (*monkey K*)), although for *monkey H* neurons in ACC, PFC, and CD were more likely to encode variables at the second order

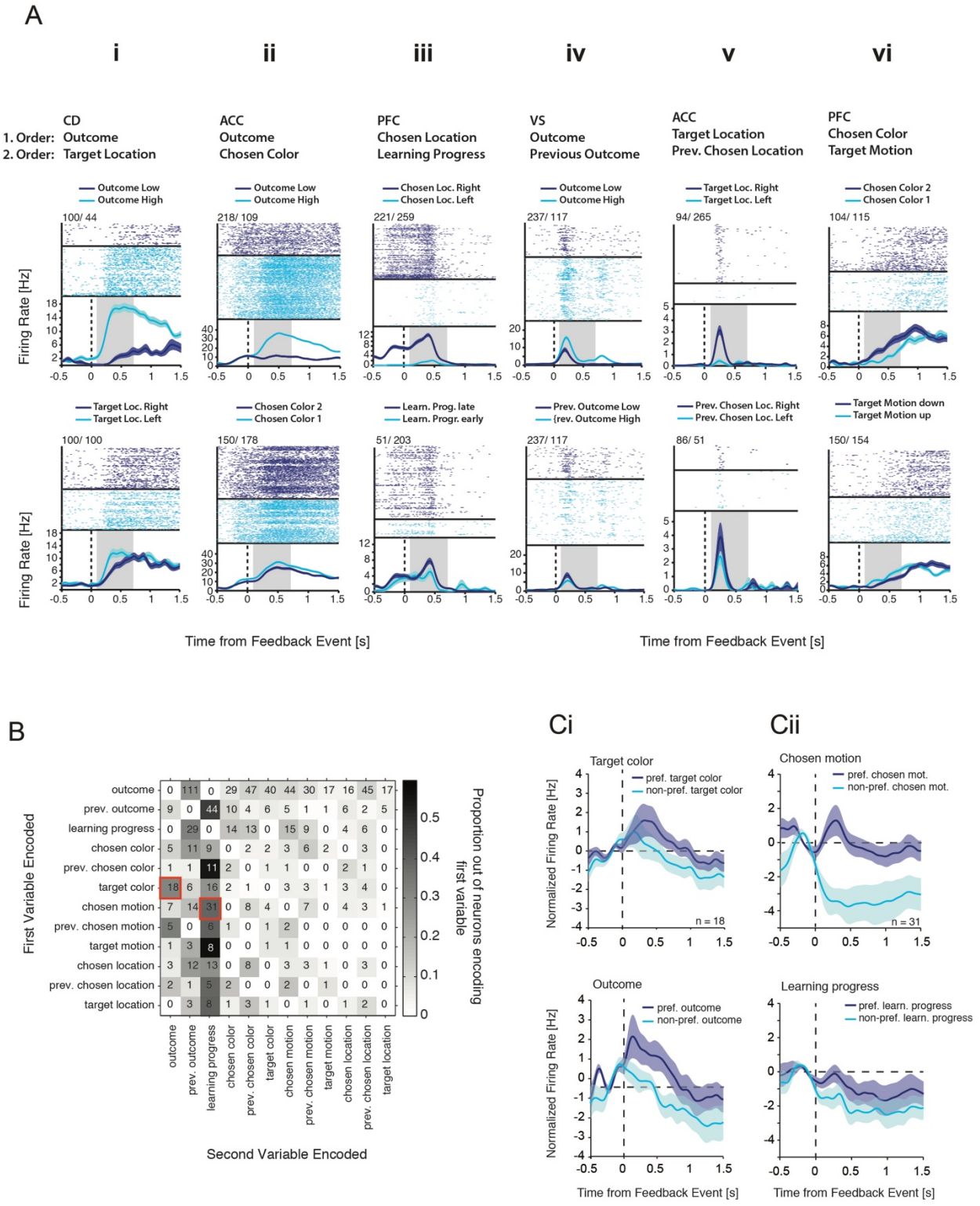

Figure 3.3 Second order encoding of task variables.

(A) Single neuron examples of neurons from different areas encoding different variables at the first and second order. Neurons i, ii, iii, and vi were recorded in *monkey H*, neurons iv and v were recorded in

monkey K. Shaded error bars represent SEMs. Grey shaded backgrounds indicate the time window of analysis. **(B)** Proportion of neurons that encoded a second task variable relative to the first task variable encoded at the time of outcome combined for *monkey H* and *K*. Red squares indicate variable combinations for which average firing rates are shown in **C**. **(C)** Average spike rates for neurons encoding target color as the first variable and outcome as the second variable (**Ci**), and neurons encoding chosen motion as the first variable and learning progress as the second variable (**Cii**), each combined for *monkey H* and *K*.

than neurons in VS (pairwise chi-square comparisons, Bonferroni-Holm multiple comparison corrected, $p < 0.05$, all others $p > 0.15$).

We find that across the range of variables encoded at the first order (Figure 3.3A), the major variables that were multiplexed are linked to the encoding of previous reward outcomes and performance. Specifically, the most common variables encoded at the second order were previous outcome (*monkey H:* 21.0%, *monkey K:* 22.8%), and learning progress (*monkey H:* 17.2%, *monkey K:* 16.7%) (Figure 3.3A, B), with both of these more often encoded at the second order than expected based on an equal distribution across all twelve variables, which was not the case for any of the other variables (Binomial test, $p < 0.001$ for both monkeys). These results were similar across areas for *monkey H*, with previous outcome encoded at the second order in 20.8%, 17.7%, 23.3% and 27.0% of variables in ACC, PFC, CD, and VS, respectively, and with learning progress encoded at the second order in 15.3%, 20.7%, of variables in ACC and PFC respectively (Binomial test, all $p < 0.01$). In PFC chosen motion was also encoded at the second order in 15.7% of variables ($p < 0.01$). For *monkey K*, split by areas, these results held for previous outcome encoded at the second order in 30.3%, 18.9%, and 27.5% of variables in ACC, CD and VS, and with learning progress encoded at the second order in 23.8% of variables in PFC (Binomial test, all $p < 0.05$). The previous suggests that a large proportion of neurons in fronto-striatal areas multiplexes information of different task variables, in particular variables that track performance over time (Figure 3.3B, combined across monkeys). The average firing rates for neurons encoding target color at the first order and outcome at the second order, as well as chosen motion at the first order and learning progress at the second order are shown combined for *monkey H* and *K* in Figure 3.3Ci and Cii, respectively. The frequency of multiplexing we found in neurons across areas, and the finding that variables that were

multiplexed at the second order tended to reflect previous outcomes and learning progress, suggests that neurons in these areas might also multiplex prediction error signals with feature-choice signals, which we tested next.

3.4.4 RL model fit

The Bayes-RL learning model was fit well to both monkeys (log likelihoods for *monkey H* and *monkey K* were 0.47 and 0.52, respectively). The model parameters best explaining the data for *monkey H*/*K* had a similar pattern with η (learning rate) = 0.22/0.25, β (selection noise) = 3.55/2.79, ϕ (dimension weighting of feature representation) = 0.68/0.98 and ω (value decay for nonchosen feature) = 0.92/0.68. These results resonate well with previous studies using a similar model architecture (Wilson and Niv, 2012; Niv et al., 2015; Hassani et al., 2017; Leong et al., 2017) (Figure 3.4A).

3.4.5 Non-specific prediction errors in fronto-striatal areas

At first, across all brain regions, we identified neurons that encoded *non-specifically* a positive prediction error (denoted as pPE in figures and text brackets), a negative prediction error (denoted as nPE in figures and text brackets) or an unsigned prediction error signal (denoted as |PE| in figures and text brackets), by correlating their firing rate around the outcome event with a trial-by-trial PE identified using an RL model (Figure 3.4A; for details *see* Methods). Unsigned prediction errors indicate the deviation from expectation independent of the valence of the outcome that can therefore be described as general surprise signals and are thought to support learning (Pearce and Hall, 1980; Courville et al., 2006). We found that overall 44.2% of neurons encoded a non-specific PE signal (*monkey H*: 42%, *monkey K*: 47.4%). Indeed, every area encoded significant amounts of all three signal types in both monkeys (one-sided binomial test, all $p < 0.05$), except for ACC in *monkey K*, which did not encode substantial amounts of negative PEs ($p > 0.05$) (Figure 3.5A). Proportions for this analysis are computed out of all neurons recorded that met the initial trial selection criteria for analysis. Positive, negative and unsigned PEs were encoded in similar

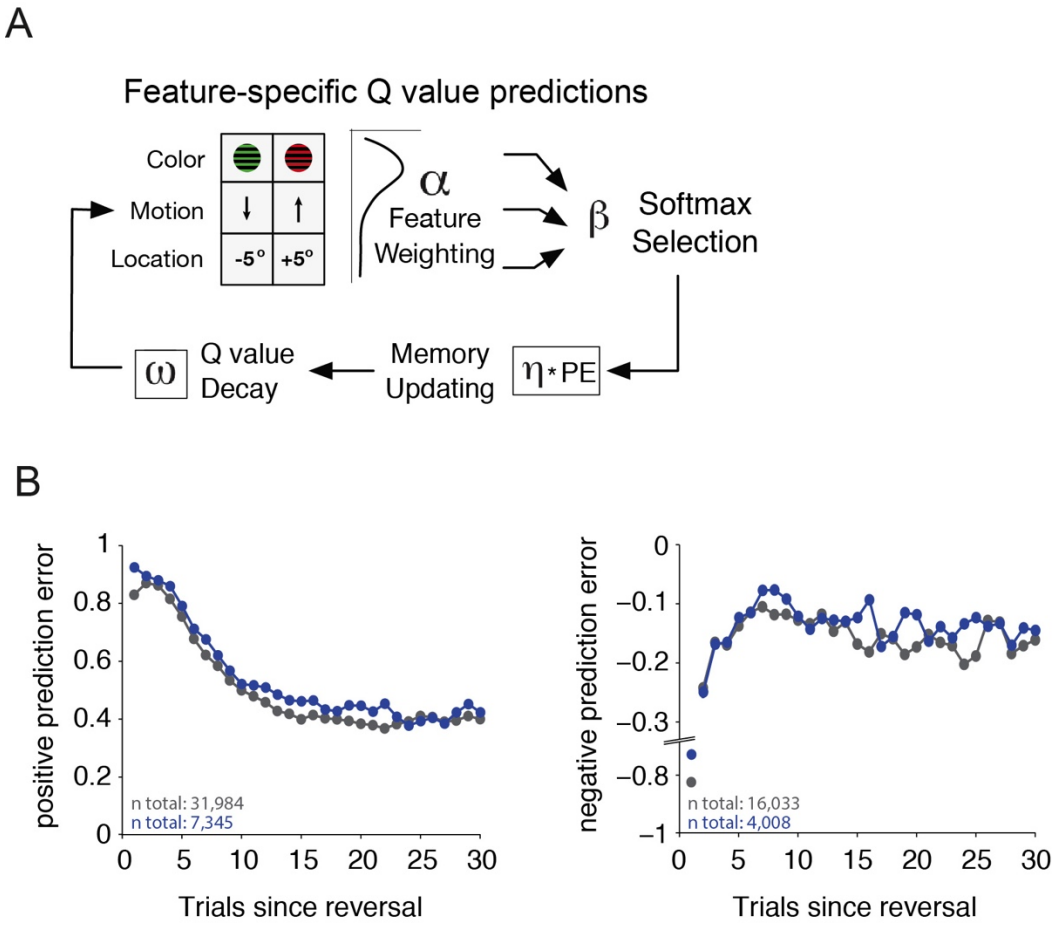

Figure 3.4 Reinforcement learning model.

(A) RL model schema. **(B)** Average block-aligned positive (left) and negative (right) prediction errors computed by the model for *monkey H* (grey) and *monkey K* (blue).

amounts in all areas (pairwise chi-square comparisons, Bonferroni-Holm multiple comparison corrected, all $p > 0.2$). As can be seen in Figure 3.5A, across areas negative PEs were less often observed than positive or unsigned PEs in an average of 12.7% of neurons (*monkey H*: 13.6%, *monkey K*: 10.6%), compared with 24.2% positive PEs (*monkey H*: 21.9%, *monkey K*: 29.7%) and 25.1% unsigned PEs (*monkey H*: 24.7%, *monkey K*: 26%) (pairwise chi-square comparison, Bonferroni-Holm multiple comparison corrected, $p < 0.001$). Notably, we found that multiple neurons encoded more than just one of

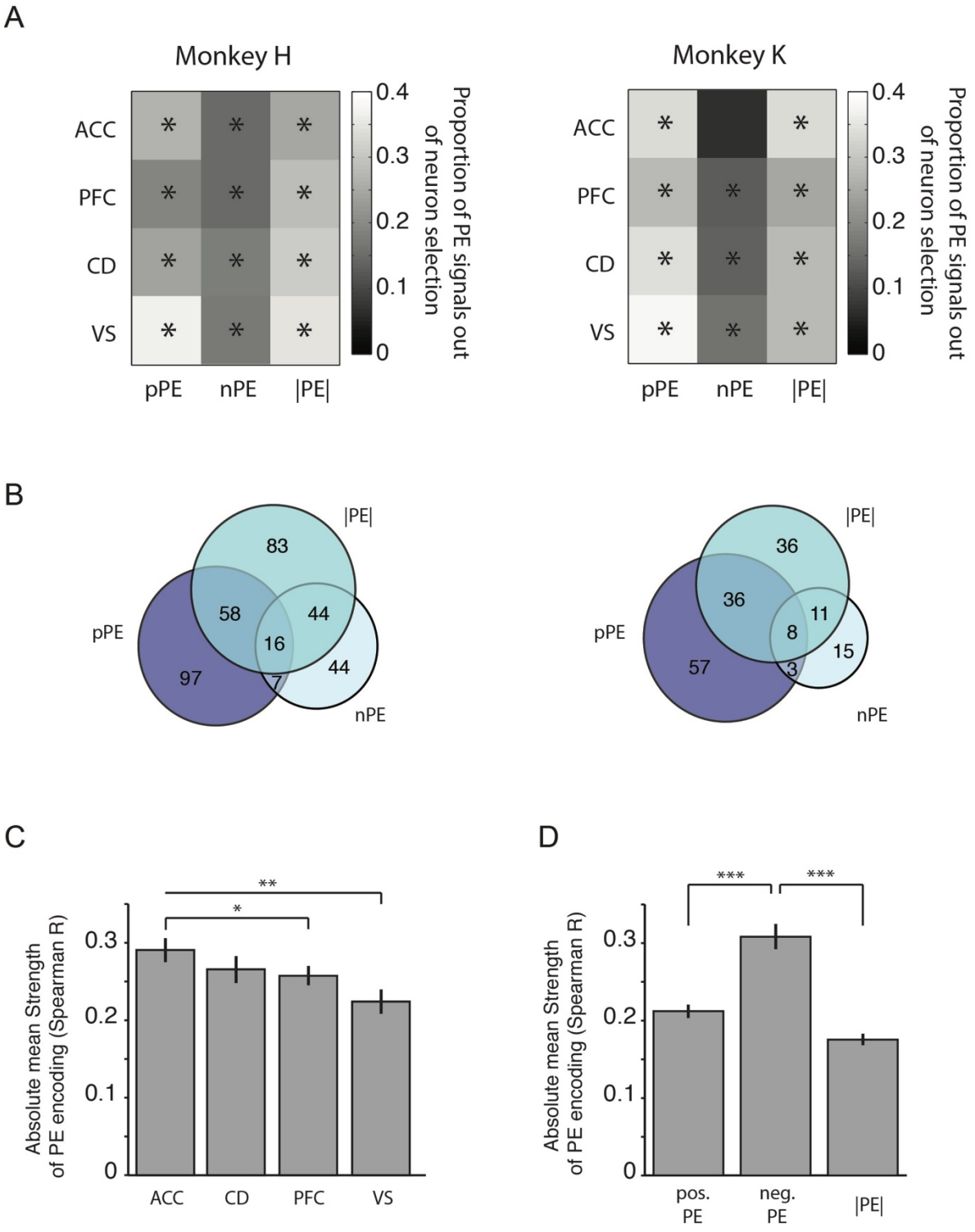

Figure 3.5 Non-specific prediction error encoding.

(A) Matrices of the proportion of non-specific PE signals observed in the four different areas and for the different PE types relative to all neurons that matched the initial trial selection for analysis for *monkey H*

(left) and *monkey K* (right). Asterisks indicate significantly larger proportions than expected by chance (binomial test, $p < 0.05$). **(B)** Overlap of neurons encoding more than one non-specific PE type (positive, negative or unsigned) for *monkey H* (left) and *monkey K* (right). Digits indicate numbers of neurons in each selection. **(C)** and **(D)**: Effect size (correlation strength) of non-specific PEs and spike rate (Spearman's R) averaged for the identified PE encoding time windows (*see* Methods for details). Neurons in ACC showed the greatest effect size of correlating with PEs followed by CD, PFC, and VS neurons (**C**). Negative PEs were on average encoded more strongly than positive or unsigned PEs (**D**). Plots in **C** and **D** are combined for *monkey H* and *K* due to the similarity of results (see Methods for values separated by monkeys). *p<0.05, **p<0.01, ***p<0.001.

the three PE types (e.g. positive and negative PE). Across areas, the proportion of these neurons was similar for both *monkey H* and *monkey K* (pairwise chi-square comparison, Bonferroni-Holm multiple comparison corrected, all $p > 0.05$). The overlap in populations encoding one versus two versus three types of PE signals is displayed in Venn diagrams in Figure 3.5B collapsed across areas. Overall, 0.9% of neurons encoded a positive and a negative PE signal (*monkey H*: 0.9%, *monkey K*: 0.9%), 8.1% encoded a positive and an unsigned PE (*monkey H*: 7.1%, *monkey K*: 10.3%), 4.7% encoded and negative and an unsigned PE (*monkey H*: 5.4%, *monkey K*: 3.1%), and 2.1% encoded all three types of PEs (*monkey H*: 2%, *monkey K*: 2.3%).

We next tested whether the effect size of PE encoding (strength of correlation) differed between areas and PE types. To do so, we identified for each neuron the time window/s in which it encoded a PE signal (for details *see* Methods) and computed a 2-way ANOVA with the parameters area (ACC, PFC, CD, VS) and PE type (positive, negative, unsigned). We found a significant main effect for both parameters ($p < 0.001$), but no interaction between them ($p > 0.05$) in both *monkey H* and *monkey K*. Post-hoc comparisons revealed that ACC showed the largest PE effect size (mean variance explained: 8.4%), followed by CD (mean variance explained: 7.3%) and PFC (mean variance explained: 6.8%), followed by VS (mean variance explained: 4.8%), with ACC in *monkey H*, and ACC, PFC and CD in *monkey K* having a significantly stronger effect size than VS (pairwise t-tests, Bonferroni-Holm multiple comparison corrected; *monkey H*: $p_{ACC\text{-}VS} = 0.0034$, *monkey K*: $p_{ACC\text{-}VS} = 0.002$, $p_{PFC\text{-}VS} = 0.002$, $p_{CD\text{-}VS} = 0.01$) (Figure 3.5C). Post-hoc comparisons also revealed that negative PE signals (mean variance

explained: 9.6%) showed a stronger mean correlation strength than positive (mean variance explained: 4.4%) and unsigned PE signals (mean variance explained: 3.2%) in both *monkey H* and *K* (pairwise t-test, multiple comparison corrected: $p < 0.001$) (Figure 3.5D). Because the effect size differences across areas and PE types was the same in *monkey H* and *K*, we combined data across the two in Figure 3.5C and D and this combined comparison additionally separated ACC from PFC, and PFC and CD from VS (Figure 3.5C) (pairwise t-tests, Bonferroni-Holm multiple comparison corrected: $p_{ACC\text{-}PFC} = 0.006$, $p_{ACC\text{-}CD} = 0.08$, $p_{ACC\text{-}VS} < 0.001$, $p_{PFC\text{-}CD} = 0.46$, $p_{PFC\text{-}VS} = 0.006$, $p_{CD\text{-}VS} = 0.004$).

3.4.6 Feature-specific prediction errors in fronto-striatal areas

We set out to determine whether neurons in fronto-striatal areas would multiplex prediction error information with feature choice information. We indeed find many neurons that encode a prediction error selectively for one chosen feature value compared with the second feature value (Figure 3.6). For instance, the VS neuron in Figure 3.6i scales its firing rate with the positive PE value when color 1 was selected in the preceding choice (top), but not when color 2 was selected in the preceding choice (middle). The PFC neuron in Figure 3.6iv scales its firing rate with the surprise signal (unsigned PE), when the stimulus selected in the preceding choice was located on the right (top), but not when it was located on the left (middle). And the ACC neuron in Figure 3.6v scales its firing rate with the negative PE (greater firing with more negative PE) when the selected stimulus in the preceding choice was moving upwards (top), but not when it was moving downwards (middle). Overall, we find that 56.3% of neurons (*monkey H*: 55.4%, *monkey K*: 58.3%) across the fronto-striatal areas tested here encoded feature-specific positive, negative and unsigned PE signals (Figure 3.7A, C). Independent of the specific feature, the proportion of neurons that encoded feature-specific positive, negative and unsigned PEs was greater than expected by chance in all areas in both monkeys, with the exception of CD in *monkey K*, which did not encode substantial amounts of feature-specific negative PEs (one-sided binomial test, all other $p < 0.05$) (Figure 3.7A, C). Similar to non-specific PE encoding, independent of the area, feature-specific unsigned PEs

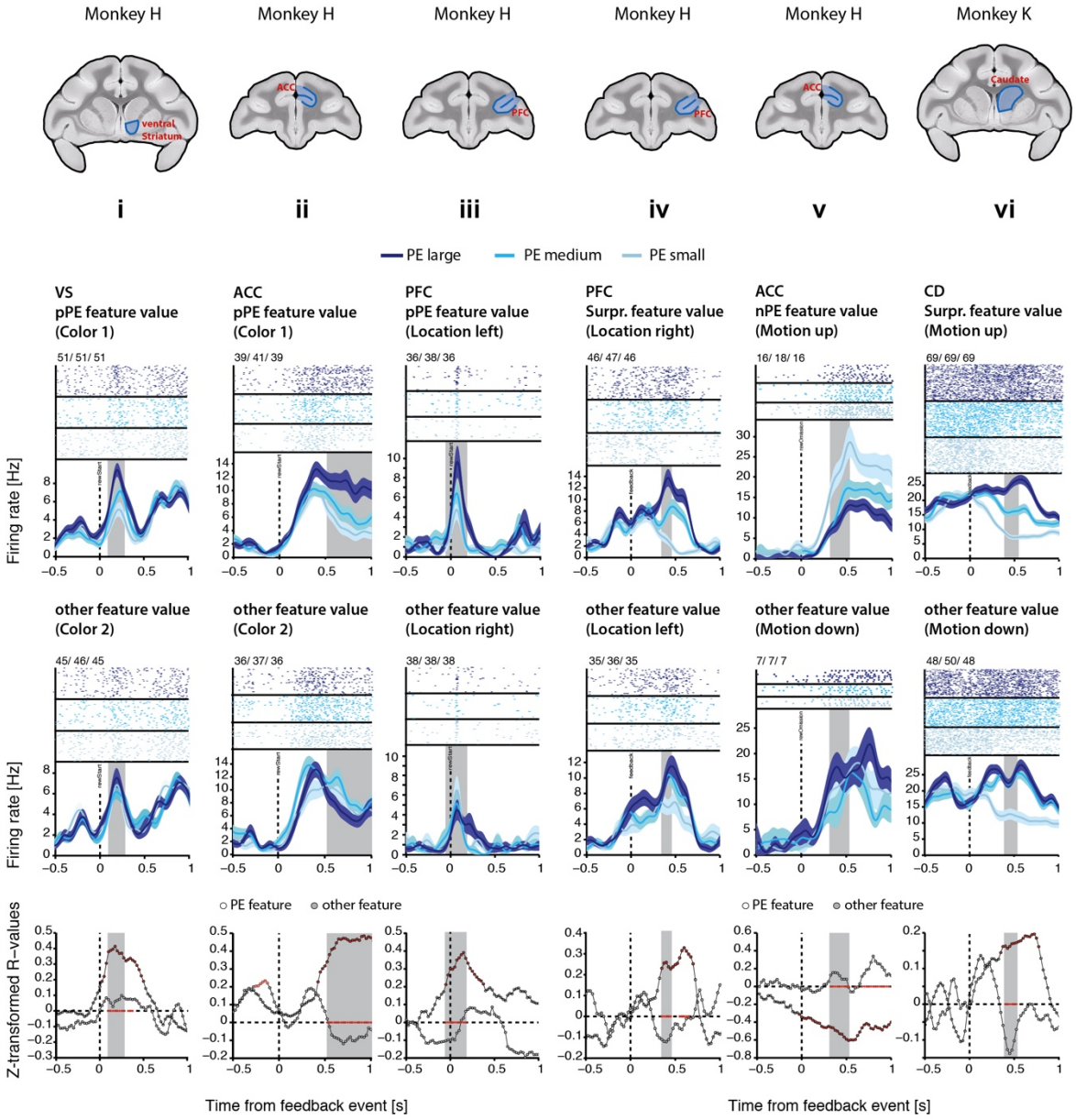

Figure 3.6 Examples of neurons encoding feature-specific PE signals.

Displayed are 6 example neurons (**i - vi**) encoding PE signals for different feature- and PE-types and from different areas and monkeys. For each neuron, the spike rasters and spike density functions are displayed for differently sized PE values (RPE large, RPE medium, RPE small) for the feature value for which a PE was encoded (top row), and for the feature value for which a PE was not encoded (middle row, 'other feature value'). The bottom row displays the z-transformed R values of the correlation between spike rate and PE for the two feature values above. The title above each column of figures indicates the area that neuron was recorded from as well as the type of feature and PE signal encoded by that neuron. Anatomical images at the top-most additionally illustrate the recording locations. From left to right, the first 5 neurons were recorded from *monkey H* and the last one (right-most) from *monkey K* as indicated above the anatomical images. Shaded error bars represent SEM.

were more prevalent than feature-specific positive PEs which were more prevalent than feature-specific negative PEs (pairwise Chi-square comparison, Bonferroni-Holm multiple comparison corrected, all $p < 0.05$).

Feature-tuning of prediction errors

When we split feature-specific PEs by the features they were specific for (color, location or motion), to determine which areas encoded which feature-specific PEs more often than expected by chance (one-sided binomial test), we revealed the following (Figure 3.7B, D): i) all areas in both monkeys selectively encoded color-specific positive PEs; ii) the most commonly encoded negative PE was a color-specific negative PE (in ACC, CD and VS in *monkey H,* and in PFC and VS in *monkey K*); iii) VS is the only area that consistently encoded color-specific negative PEs in both monkeys; iv) PFC is the only area that consistently encoded positive PEs for all feature types in both monkeys; and v) all areas in both monkeys encoded color-, location- and motion-specific unsigned PEs, with the exception of VS in *monkey K*, which did not encode substantial amounts of location-specific unsigned PEs (Figure 3.7B, D). Pairwise comparisons between areas showed that ACC and CD encoded significantly more location-specific unsigned PEs than VS in *monkey K* (pairwise chi-square comparisons, Bonferroni-Holm multiple comparison corrected, $p < 0.05$). To illustrate differences in the prevalence of feature-specific PEs (combined across monkeys), we normalized prevalence (proportion) across areas, features, and PE types to scale between 0.1 and 1 in Figure S3A Appendix A. Rank-ordering the first five signals across all signals showed that the three feature-specific unsigned PEs in ACC were most prevalent across all feature-specific PE signals, followed by motion-specific unsigned PEs in PFC and location-specific unsigned PEs in CD (Figure S3A right Appendix A).

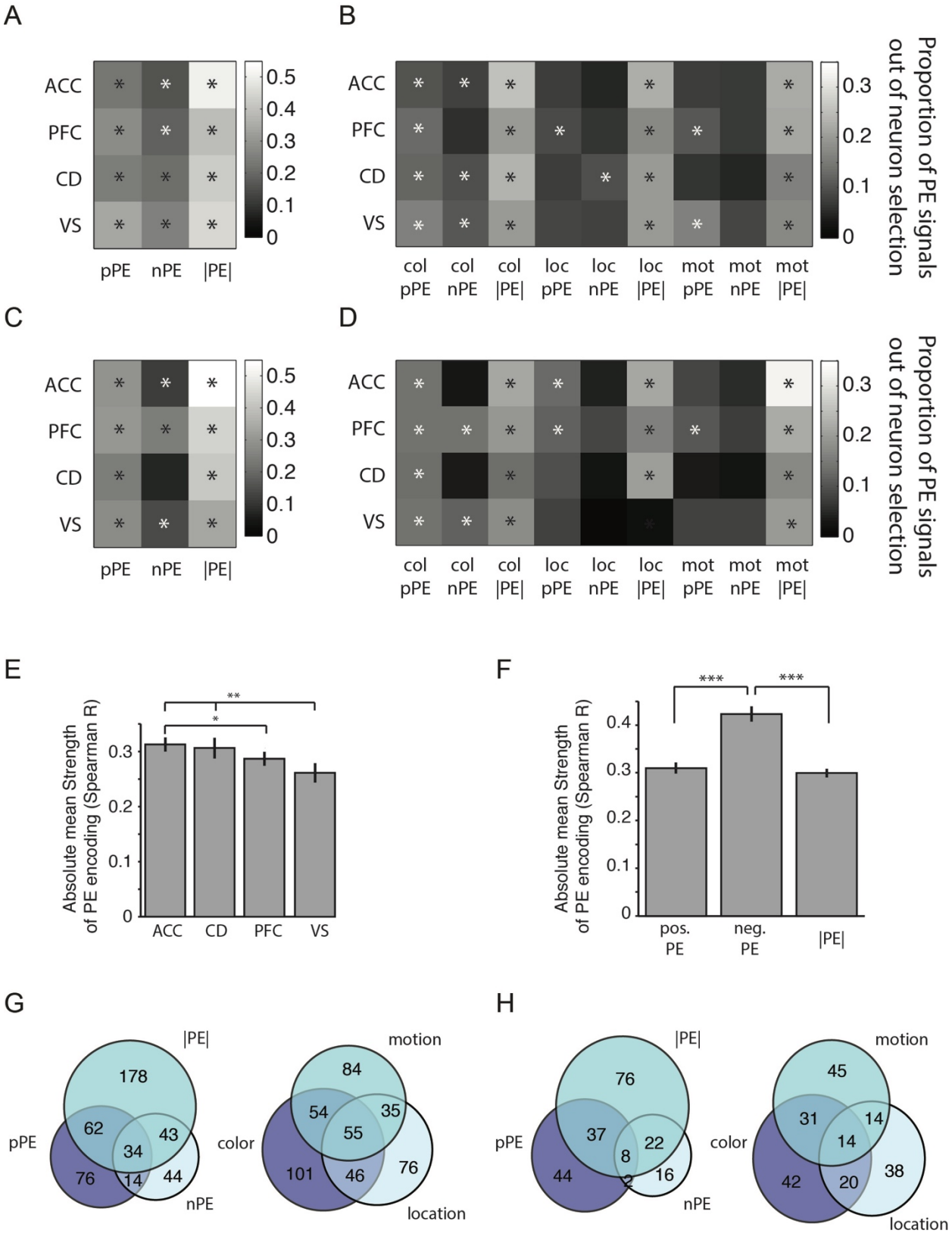

Figure 3.7 Feature-specific prediction error encoding.

(A) and **(C)** show matrices of the proportion of feature-specific PE signals observed in the four different areas for the three different PE types relative to neurons that matched the initial trial selection for analysis in *monkey H*, and *monkey K*, respectively. **(B)** and **(D)** show matrices of the proportion of feature-specific PE signals observed for each area and PE type separated by the specific features that were encoded (color, location or motion), for *monkey H* and *K*, respectively. Asterisks indicate significantly larger proportions than expected by chance (binomial test, $p < 0.05$). **(E)** and **(F)** Overlap of neurons encoding a PE signal for more than one PE type (positive, negative or unsigned) (left) or for more than one PE feature (right) for *monkey H* and *K*, respectively. **(G)** and **(H)** Effect size (correlation strength) of feature-specific PEs and spike rate (Spearman's R) averaged for the identified PE encoding time windows. Neurons in ACC showed the greatest effect size of correlating with PEs followed by CD, PFC, and VS neurons (**G**). Negative PEs were on average encoded more strongly than positive or unsigned PEs (**H**). Plots in **G** and **H** are combined for *monkey H* and *K* due to the similarity of results (see Methods for values separated by monkeys). *p<0.05, **p<0.01, ***p<0.001.

Across PE types and areas, the specific feature values that were represented in feature-specific PEs were distributed evenly in *monkey H* (49.8% color 1- red, 50.1% color 2 – green; 50.2% location 1 – left, 49.8% location 2 – right, 52.5% motion 1 – up, 47.5% motion 2 – down) (binomial tests, all $p > 0.4$), while *monkey K* had a slight bias for encoding PEs specific for color 2 (37% color 1- yellow, 63% color 2 – cyan; 52.9% location 1 – left, 47.1% location 2 – right, 41.5% motion 1 – up, 58.5% motion 2 – down) (binomial tests, $p_{color} = 0.004$, $p_{location} = 0.62$, $p_{motion} = 0.08$).

Effect sizes of feature-specific prediction errors

To test whether the effect sizes (mean correlation strength) of feature-specific PE encoding varied with PE type, area, or feature type, we determined for every neuron the time window/s in which it encoded a PE signal (for details *see* Methods) and computed a 3-way ANOVA with the parameters PE type (positive, negative, unsigned), area (ACC, PFC, CD, VS), and feature type (color, location, motion). We found main effects for the parameters PE type and area and no main effect for feature type in both monkeys (3-way ANOVA, $p < 0.05$). Similar to non-specific PE encoding, post-hoc comparisons revealed that ACC (mean variance explained: 9%) showed the largest feature-specific PE effect size, followed by CD (mean variance explained: 8.5%) and PFC (mean variance explained: 7.6%), followed by VS (mean variance explained: 6.1%), with ACC in *monkey H*, and ACC, PFC, and CD as a trend in

monkey K having a significantly stronger effect size than VS (pairwise t-tests, Bonferroni-Holm multiple comparison corrected; *monkey H*: $p_{ACC\text{-}VS} = 0.0034$, *monkey K*: $p_{ACC\text{-}VS} = 0.004$, $p_{PFC\text{-}VS} = 0.027$, $p_{CD\text{-}VS} = 0.06$) (Figure 3.7E combined across monkeys). Post-hoc comparisons also revealed that negative feature-specific PE signals showed a stronger mean correlation strength than positive and unsigned feature-specific PE signals in both *monkey H* and *K* (pairwise t-test, multiple comparison corrected: $p < 0.001$) (Figure 3.7F). Notably, we found that independent of the previous parameters, feature-specific PEs (mean variance explained: 10.9%) had a greater effect size, and were therefore more strongly encoded than non-specific PEs (mean variance explained: 4.8%) in both *monkey H* and *K* (t-test, both $p < 0.001$). Finally, out of all units that encoded a PE signal, 83% encoded at least one PE signal feature-specifically, highlighting that feature-specific PE encoding is remarkably prevalent. To illustrate differences in the effect sizes of feature-specific PEs across the two monkeys, we normalized effect size across areas, features, and PE types to scale between 0.1 and 1 (Figure S3B Appendix A). Rank-ordering the first five signals across all signals showed that location and motion-specific negative PEs in CD showed the greatest effect size, followed by all feature-specific negative PEs in ACC (Figure S3B right Appendix A). To illustrate which feature-specific PE signals occurred most prevalently and showed the greatest effect size, we multiplied normalized prevalence with normalized effect size in Figure S3C Appendix A. This computation highlights all feature-specific unsigned PEs (surprise signals) in ACC, followed by location-specific unsigned PEs in CD and color-specific negative PEs in VS (Figure S3C right Appendix A).

Overlap and anatomical locations of feature-specific prediction error units

We observed that some neurons encoded more than just one feature-specific PE signal. The proportion of those neurons that encoded more than one PE type (positive, negative, unsigned) or more than one feature type (color, location, motion) was similar across areas in both monkeys (pairwise chi-square comparisons, Bonferroni-Holm multiple comparison corrected, all $p > 0.1$). Overall, 1.3% of neurons encoded a positive and negative feature-specific PE (*monkey H*: 1.7%, *monkey K*: 0.6%), 8.5% encoded a positive

and unsigned feature-specific PE (*monkey H*: 7.6%, *monkey K*: 10.6%), 5.6% encoded a negative and an unsigned feature-specific PE (*monkey H*: 5.3%, *monkey K*: 6.3%), and 3.6% encoded a feature-specific PE for all three PE types (*monkey H*: 4.2%, *monkey K*: 2.3%) (Figure 3.7G, H left). Furthermore, 5.7% of neurons encoded a color- and a location-specific PE (*monkey H*: 5.6%, *monkey K*: 5.7%), 7.3% of neurons encoded a color- and motion-specific PE (*monkey H*: 6.6%, *monkey K*: 8.9%), 4.2% of neurons encoded a location- and motion-specific PE (*monkey H*: 4.3%, *monkey K*: 4%), and 5.9% of neurons encoded a PE for all three features (*monkey H*: 6.7%, *monkey K*: 4%) (Figure 3.7G, H right).

Anatomical distributions within an area of feature-specific PE neurons showed no apparent clustering according to type, feature specificity or feature identity. Highlighted in Figures S3 – S6 are the anatomical locations of all feature-specific PE neurons in caudate and ventral striatum (Figure S4, S6 Appendix A), and in ACC and PFC (Figure S5, S7 Appendix A) for *monkey H* and *K*. Neurons are differentiated based on the feature they were selective for, as well as the number and type of features they were selective for should they have been selective for more than one (Figure S4 – S7 Appendix A). Color-, location-, and motion-selective PE neurons were found similarly distributed within the areas tested.

3.4.7 Area-dependent time course of prediction error signals

We found that across the fronto-striatal areas, neurons encoded feature-specific positive, negative and unsigned PEs. This allowed to test whether there was a specific temporal ordering of the emergence of PE information across areas. The answer to this question could potentially allow some insight into where these signals originated. In all brain regions, PE encoding was evident in single cells as early as at the time of the outcome event and as late as 1500ms following the outcome event. We quantified the time courses of PE encoding by identifying for each neuron the full time window in which it encoded a PE signal; all the time bins in such a time window were then entered into the time course analysis. Across

single neurons, independent of the PE type, the time of PE encoding occurred earlier for non-specific (617 ± 1ms) compared with feature-specific PEs (739 ± 1ms) in *monkey H* and *monkey K* (t-test, both $p < 0.0001$), and this was the case in all areas (t-test, all $p < 0.001$). Within a given area we could then compute and compare the cumulative sums of PE encoding across neurons for different PE types (*see* Supplementary Results, Appendix B). Differences in the cumulative sums alone do not yet allow a judgement in which area PE signals emerged sooner; we therefore additionally determined for each area the time point at which its cumulative sum reached a 25% threshold and compared the time of this threshold to determine in which area PE signals emerged soonest. Similar to single neurons, the time point at which 25% of non-specific PEs were encoded across areas (293 ± 3ms) occurred sooner than the time point at which 25% of feature-specific PEs were encoded (384 ± 2ms), independent of the type of PE, in *monkey H* (t-test, $p = 0.016$) and as a trend in *monkey K* (t-test, $p = 0.073$). The previous suggests that feature-specific PE signals were generally encoded later than non-specific PE signals.

For a comparison between areas, we only analyzed the time courses of PE signals that were encoded in statistically reliable populations in both monkeys and in at least two areas (Figure 3.7B, D). This restricted the analyses to color-, location-, and motion-specific unsigned PEs, color-specific positive and negative PEs, in addition to nonspecific positive, negative and unsigned PEs. Feature-specific unsigned PEs were combined across feature-types due to the qualitative similarity between them with regards to the order of areas in which they emerged first (see Figures S7/S8 D-F for a split by features). Figure 3.8 summarizes the time points at which 25% of PE encoding occurred across areas and different PE signals, which are described in detail in the following and in Figure 3.9, S7 and S8. In general, the temporal dynamics of area-specific PE encoding differed quite substantially between the two monkeys and they will therefore be discussed separately in the following.

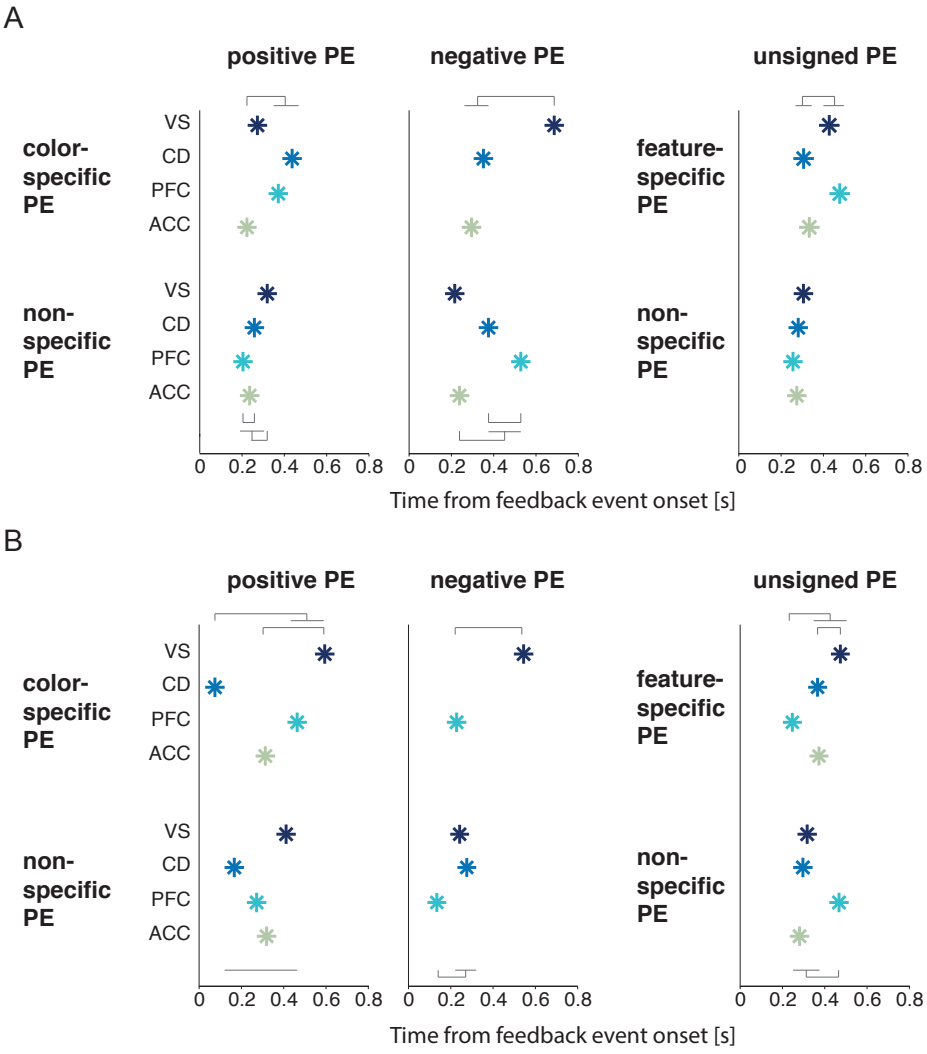

Figure 3.8 Summary of PE time course encoding.

Summarized in one comprehensive plot are the time points at which 25% of PE encoding occurs as a measure of PE latency (25% of cumulative sums) shown in detail in Figure 3.9, 3.10, S7 and S8. **(A)** The top row shows the time of 25% PE encoding for color-specific positive (left), color-specific negative (middle) and feature-specific unsigned PEs (right, pooled across feature types) for each area in *monkey H*. The bottom row shows the time of 25% PE encoding for non-specific positive (left), negative (middle), and unsigned (right) PEs. Note that if an area did not encode a given PE signal in a large enough population it is omitted from that specific time course analysis. **(B)** same as **A** for *monkey K*.

In *monkey H*, we found that the time at which 25% of *color-specific positive* PEs was encoded occurred sooner in ACC (223 ms) and VS (273 ms) compared with PFC (373 ms) and CD (438 ms)

(Randomization statistic: $p_{(ACC-PFC)} = 0.024$; $p_{(ACC-CD)} < 0.001$; $p_{(ACC-VS)} = 0.172$; $p_{(PFC-CD)} = 0.562$; $p_{(PFC-VS)} = 0.084$; $p_{(CD-VS)} = 0.002$) (Figure 3.9A inset at the bottom, Figure 3.8A left) (*see* Supplementary Results, Appendix B, for statistical differences of cumulative sums). Notably, the time course of encoding of color-specific positive PEs differed from the time course of encoding of non-specific positive PEs. The threshold of 25% *non-specific positive* PE encoding was reached soonest by PFC (205 ms) followed by ACC (238 ms) and CD (260 ms) and then VS (320 ms) (Randomization statistic: $p_{(ACC-PFC)} = 0.102$; $p_{(ACC-CD)} = 0.358$; $p_{(ACC-VS)} < 0.001$; $p_{(PFC-CD)} = 0.018$; $p_{(PFC-VS)} < 0.001$; $p_{(CD-VS)} = 0.014$; Figure S8A Appendix A, Figure 3.8A left).

Color-specific negative PEs were encoded sooner in ACC (295 ms) and CD (353 ms), followed by VS (688 ms) (Randomization statistic: $p_{(ACC-VS)} < 0.001$; $p_{(CD-VS)} < 0.001$, all others $p > 0.2$) (Figure 3.9B inset at the bottom, Figure 3.8A middle). In contrast, *non-specific negative* PEs were encoded soonest in VS (218 ms) and ACC (240 ms), followed by CD (378 ms), followed by PFC (530 ms) (Randomization statistic: $p_{(ACC-PFC)} < 0.001$; $p_{(ACC-CD)} = 0.024$; $p_{(ACC-VS)} = 0.604$; $p_{(PFC-CD)} = 0.018$; $p_{(PFC-VS)} < 0.001$; $p_{(CD-VS)} = 0.026$) (Figure S8B Appendix A, Figure 3.8A middle).

Feature-specific unsigned PEs were encoded soonest in CD (310 ms) and ACC (335 ms), followed by VS (430 ms) and PFC (480 ms) (Randomization statistic: $p_{(ACC-PFC)} < 0.001$; $p_{(ACC-CD)} = 0.508$; $p_{(ACC-VS)} = 0.001$; $p_{(PFC-CD)} < 0.001$; $p_{(PFC-VS)} = 0.08$; $p_{(CD-VS)} = 0.002$) (Figure 3.9C inset at the bottom, Figure 3.8A right). In contrast, *non-specific unsigned* PEs were encoded at similar times across all areas (ACC: 275ms; PFC: 258ms; CD: 283ms; VS: 308ms; Randomization statistic: all $p > 0.1$) (Figure S8C Appendix A, Figure 3.8A right). The previous suggests that in *monkey H*, area-specific time courses of feature-specific PE encoding differed based on the PE type and were generally independent of non-specific PE encoding (Figure 3.8A).

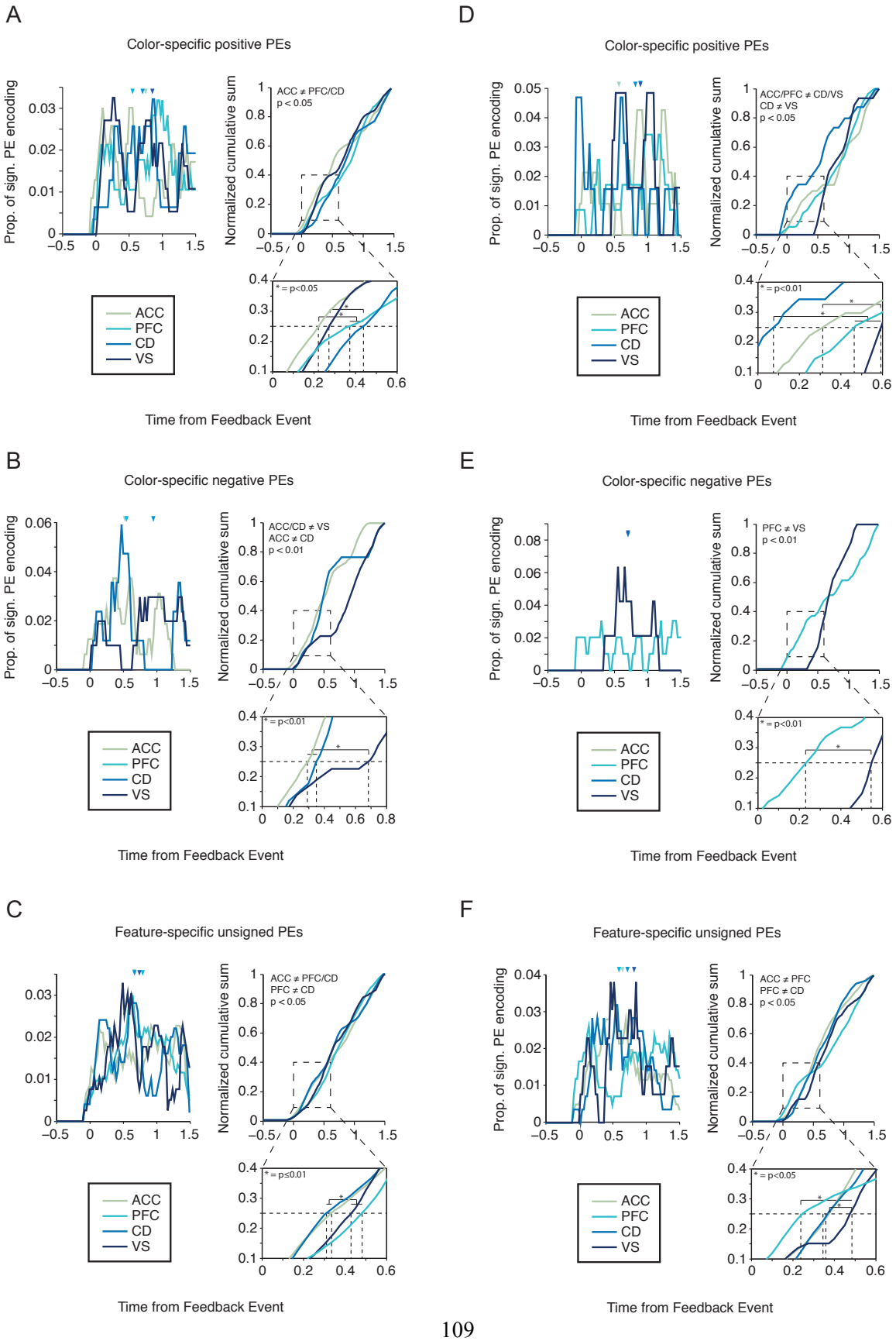
A
Color-specific positive PEs
Prop. of sign. PE encoding
Normalized cumulative sum
ACC ≠ PFC/CD
p < 0.05
* = p<0.05
ACC
PFC
CD
VS
Time from Feedback Event
D
Color-specific positive PEs
Prop. of sign. PE encoding
Normalized cumulative sum
ACC/PFC ≠ CD/VS
CD ≠ VS
p < 0.05
* = p<0.01
ACC
PFC
CD
VS
Time from Feedback Event
B
Color-specific negative PEs
Prop. of sign. PE encoding
Normalized cumulative sum
ACC/CD ≠ VS
ACC ≠ CD
p < 0.01
* = p<0.01
ACC
PFC
CD
VS
Time from Feedback Event
E
Color-specific negative PEs
Prop. of sign. PE encoding
Normalized cumulative sum
PFC ≠ VS
p < 0.01
* = p<0.01
ACC
PFC
CD
VS
Time from Feedback Event
C
Feature-specific unsigned PEs
Prop. of sign. PE encoding
Normalized cumulative sum
ACC ≠ PFC/CD
PFC ≠ CD
p < 0.05
* = p≤0.01
ACC
PFC
CD
VS
Time from Feedback Event
F
Feature-specific unsigned PEs
Prop. of sign. PE encoding
Normalized cumulative sum
ACC ≠ PFC
PFC ≠ CD
p < 0.05
* = p<0.05
ACC
PFC
CD
VS
Time from Feedback Event

Figure 3.9 Time courses of feature-specific PE signals.

(A) Time courses of color-specific positive PEs. *Left:* line histograms of the proportion of significant PE encoding time-resolved across neurons for each area. Triangles indicate means. *Right:* normalized cumulative sums of the histograms to the left. Text in the top left corner indicates for which area combination these cumulative sums differed significantly (Kolmogorov-Smirnoff test, Bonferroni-Holm multiple-comparison correction). *Bottom right:* Magnification of the cumulative sums around the 25% window. Horizontal bars with asterisks indicate significant differences in the time 25% of the cumulative sum is reached between areas (randomization procedure). **(B)** and **(C)** same as **A** for feature-specific negative and unsigned PEs, respectively. **(D) – (F),** same as **A - C** for *monkey K*.

In *monkey K*, we found that the time at which 25% of *color-specific positive* PEs were encoded occurred soonest in CD (75 ms), followed by ACC (313 ms) and PFC (465 ms), followed by VS (595 ms) (Randomization statistic: $p_{(ACC\text{-}PFC)} = 0.262$; $p_{(ACC\text{-}CD)} = 0.068$; $p_{(ACC\text{-}VS)} = 0.004$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} = 0.054$; $p_{(CD\text{-}VS)} < 0.001$) (Figure 3.9D inset at the bottom, Figure 8B left) (*see* Supplementary Results, Appendix B, for statistical differences of cumulative sums). The time course of encoding of color-specific positive PEs was similar to the time course of encoding of non-specific positive PEs. The threshold of 25% *non-specific positive* PE encoding was reached soonest by CD (168 ms) followed by PFC (273 ms) and ACC (320 ms) and then by VS (413 ms) (Randomization statistic: $p_{(PFC\text{-}ACC)} = 0.218$; $p_{(ACC\text{-}CD)} < 0.001$; $p_{(ACC\text{-}VS)} < 0.001$; $p_{(PFC\text{-}CD)} = 0.006$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} < 0.001$; Figure S9A Appendix A, Figure 3.8B left).

Color-specific negative PEs were encoded sooner in PFC (230 ms) compared with VS (548 ms (Randomization statistic: $p_{(PFC\text{-}VS)} = 0.008$) (Figure 3.9E inset at the bottom, Figure 3.8B middle). The time course of *non-specific negative* PEs was similar, whereby 25% of PE encoding occurred soonest in PFC (120 ms), followed by VS (223 ms), followed by CD (263 ms) (Randomization statistic: $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} = 0.008$; all other $p > 0.1$) (Figure S9B Appendix A, Figure 3.8B middle).

Feature-specific unsigned PEs were encoded soonest in PFC (248 ms) followed by CD (368 ms) and ACC (375 ms), followed by VS (475 ms) (Randomization statistic: $p_{(ACC\text{-}PFC)} = 0.002$; $p_{(ACC\text{-}CD)} = 0.764$; $p_{(ACC\text{-}VS)} = 0.014$; $p_{(PFC\text{-}CD)} = 0.01$; $p_{(PFC\text{-}VS)} = 0.026$; $p_{(CD\text{-}VS)} = 0.042$) (Figure 3.9F inset at the bottom, Figure 3.9B right). This time course differed for *non-specific unsigned* PEs, for which we found no

difference in the time point at which 25% of PE encoding occurred in ACC (283 ms), CD (298 ms), and VS (318 ms) followed later by PFC (468 ms) (Randomization statistic: $p_{(ACC-PFC)} = 0.002$; $p_{(PFC-CD)} < 0.001$; $p_{(PFC-VS)} = 0.086$; all other $p > 0.3$) (Figure S9C Appendix A, Figure 3.8B right).

The previous suggests that in *monkey K*, area-specific time courses of feature-specific PE encoding differed based on the PE type but were generally similar to non-specific PE encoding.

Overall, it seems that no overarching temporal ordering of PE encoding in frontal and striatal areas is apparent, but instead that it varies substantially with the type of PE signal that is encoded as well as with the individual. The most salient point to take away is that feature-specific PEs were encoded later than non-specific PEs, potentially suggesting an increased need for processing resources when information about the preceding feature choice must be integrated.

3.4.8 Time-course of trial outcome encoding versus prediction error encoding

PE encoding might be part of encoding reward outcome per se, and could thus either following reward outcome encoding, or alternatively might be computed partially independent of outcome and may thus follow a partially different time course. We therefore tested whether differences we find with regards to the timing of PE encoding between areas are related to differences in reward encoding per se between these areas. For this we computed time-resolved correlations between firing rate and trial outcome with the same parameters as for the PE analyses, for all neurons that had been identified to encode trial outcome in the initial regression analysis (Figure 3.10Ai, see for comparison average PE encoding in 10Aii). In single neurons, reward encoding was evident as early as at the time of outcome and as late as 1500ms following the time of outcome. For each neuron, we identified the time range of significant outcome encoding (equivalently to the PE analysis). Across areas, outcome was encoded slightly earlier than non-specific PE signals at 598 ± 8ms, which was significantly earlier in *monkey K* (t-test, $p < 0.001$),

and earlier as a trend in *monkey H* (t-test, p = 0.088). Reward encoding occurred sooner than feature-specific PE encoding in both animals and in all areas (t-test, all p < 0.001).

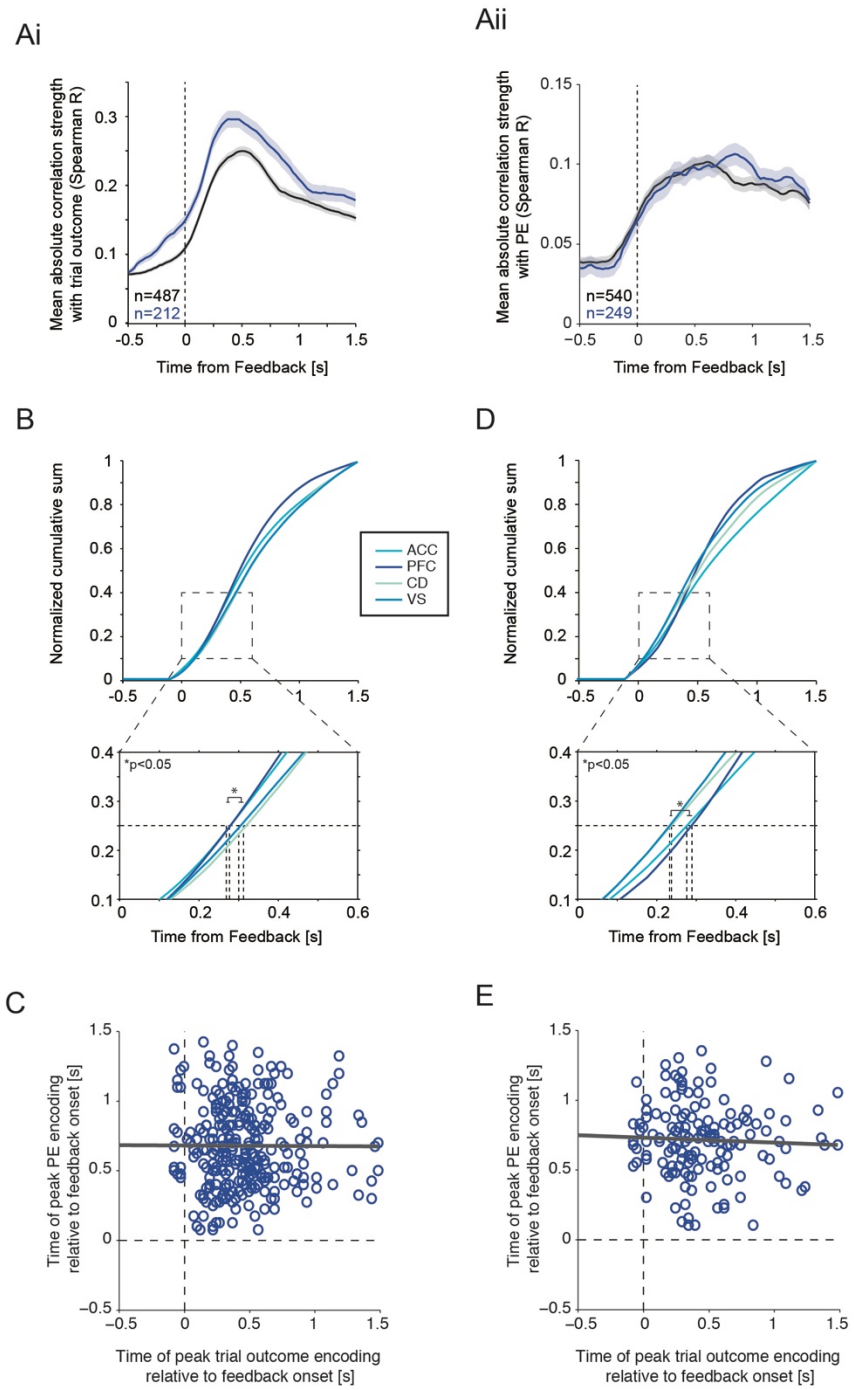

Figure 3.10 Latency comparison of PE- and reward-encoding.

(A) Time-resolved average correlation of spike rate with reward (**Ai**) and with PE (**Aii**) for *monkey H* (grey) and *monkey K* (blue) (independent of PE type). **(B)** Latency of reward encoding for all areas in *monkey H*. Conventions are the same as in Figure 7. **(C)** Correlation of the time of peak encoding of reward and the time of peak encoding of PE for those neurons that encoded both significantly for *monkey H* (Spearman correlation). **(D)** and **(E)** same as **B** and **C** for *monkey K*.

Similar to the PE analysis we computed the cumulative distributions of outcome encoding for each area. For *monkey H*, we found that 25% of trial outcome encoding occurred at the same time in ACC (278 ms) and PFC (278 ms) followed by VS (303 ms) and CD (318 ms) (Randomization statistic: $p_{(ACC\text{-}PFC)} = 0.83$; $p_{(ACC\text{-}CD)} < 0.001$; $p_{(ACC\text{-}VS)} = 0.016$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} = 0.022$; $p_{(CD\text{-}VS)} = 0.178$) (Figure 3.10B). This suggests that for *monkey H* trial outcome per se is encoded sooner in the fronto-cingulate regions compared with the striatal regions. For *monkey K*, we found that 25% of trial outcome encoding first occurred in VS (233 ms) and CD (235 ms) followed by ACC (276 ms) and PFC (288 ms) (Randomization statistic: $p_{(ACC\text{-}PFC)} = 0.412$; $p_{(ACC\text{-}CD)} = 0.008$; $p_{(ACC\text{-}VS)} = 0.012$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} = 0.836$) (Figure 3.10D). This suggests that, in contrast to *monkey H*, in *monkey K* outcome per se was encoded sooner in striatal regions compared with fronto-cingulate regions.

However, consistent across monkeys, these time courses of outcome encoding did not match the time courses of encoding of most PE signals. We next tested whether the time of the first bin that was identified as encoding a PE or reward signal, as well as the time of peak correlation with PE or reward, of those neurons that were identified as encoding a PE signal as well as a trial outcome signal, was correlated across the two measures. Neither for *monkey H*, nor for *monkey K*, were the first time of encoding or the time of peak encoding correlated between PE and reward encoding (Spearman correlation, monkey H: $R_{peak} = 0.048$, $p_{peak} = 0.93$, $R_{first} = 0.047$ $p_{first} = 0.43$; monkey K: $R_{peak} = -0.049$, $p_{peak} = 0.55$, $R_{first} = 0.004$ $p_{first} = 0.97$; see Figure 3.10C, E for correlation of peak encoding times). Therefore, although the time courses of reward encoding differed between the two animals, across all neurons, the time of PE encoding was generally independent of the time of trial outcome encoding, which

suggests that the time course differences we observe in encoding of PE signals between areas was not due to inherent time course differences in the encoding of trial outcome.

3.4.9 Cell-type specificity of prediction error signals

To test whether feature-specific and non-specific PE signals were preferentially encoded by putative interneurons or putative principal cells, we classified highly isolated PE encoding neurons in ACC and PFC into putative pyramidal cells (broad-spiking) and putative interneurons (narrow-spiking) (Figure 3.11A, B, Figure 3.12A, B), and highly isolated PE encoding neurons in CD and VS into putative medium-spiny neurons (MSNs, broad-spiking) and putative interneurons (narrow-spiking) (Figure 3.11F, G, Figure 3.12F, G). Based on the population distribution recorded in *monkey H*, we expected to observe 72% of putative pyramidal cells and 28% of putative interneurons in ACC/PFC of PE signal encoding neurons, and 73% of putative MSNs and 27% of putative interneurons in CD/VS of PE signal encoding neurons. We observed PE signals in 57% putative pyramidal cells (n = 54) and in 43% putative interneurons (n = 40) in ACC/PFC, and in 66% putative MSNs (n = 35) and in 34% putative interneurons (n = 18) in CD/VS of *monkey H*. Based on the population distribution recorded in *monkey K*, we expected to observe 82% of putative pyramidal cells and 18% of putative interneurons in ACC/PFC of PE signal encoding neurons, and 77% of putative MSNs and 23% of putative interneurons in CD/VS of PE signal encoding neurons. We observed PE signals in 73% putative pyramidal cells (n = 27) and in 27% putative interneurons (n = 10) in ACC/PFC, and in 69% putative MSNs (n = 20) and in 31% putative interneurons (n = 9) in CD/VS of *monkey K*. Due to the limited number of neurons, we pooled units into those encoding feature-specific PEs and those encoding non-specific PEs.

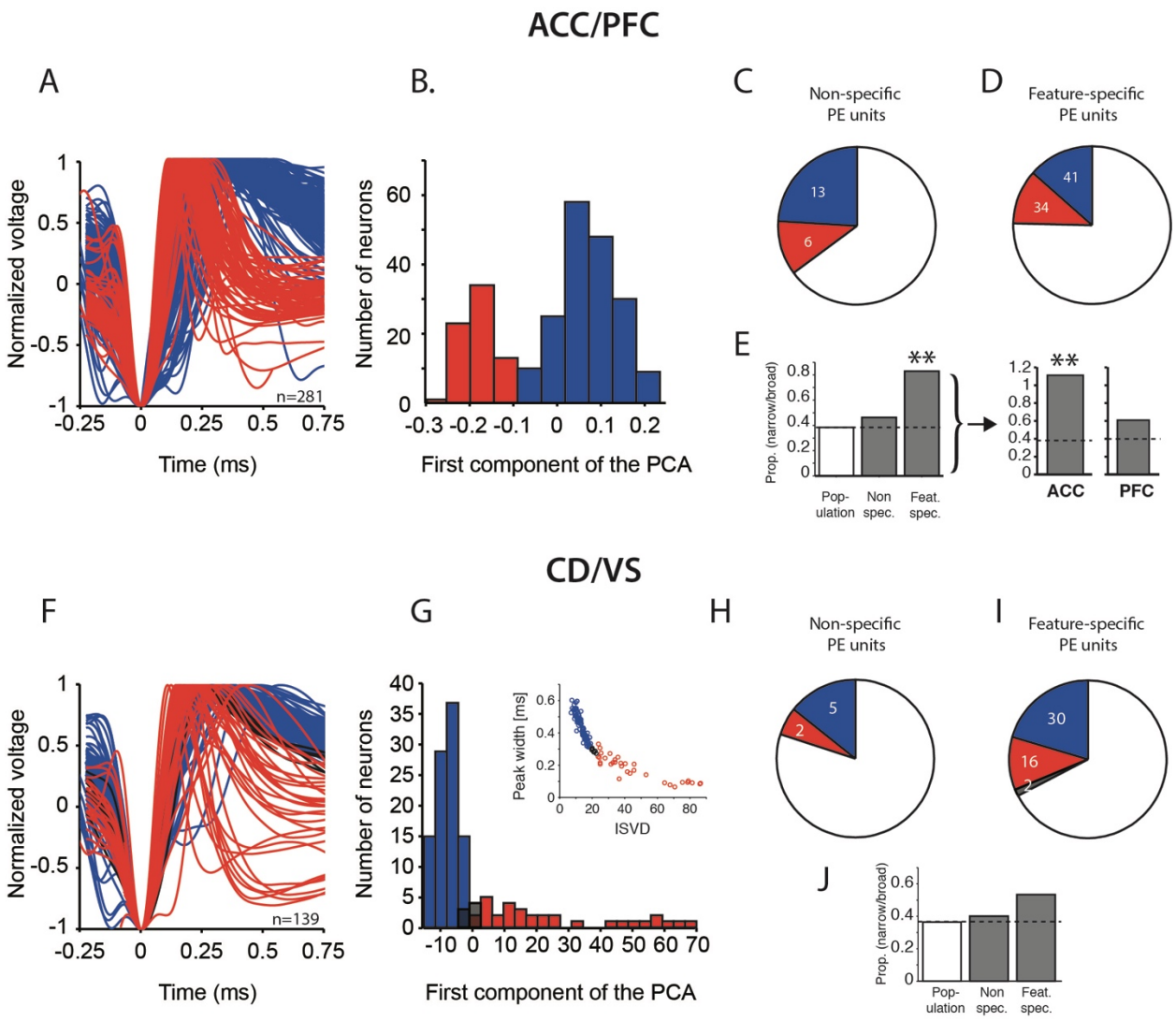

Figure 3.11 Cell-types encoding prediction errors in *monkey H*.

(A)-(E) for ACC/PFC units. **(A)** Waveforms of all highly isolated single units recorded, identified as putative interneurons (narrow-spiking, red), putative pyramidal cells (broad-spiking, blue), or unidentified (black) (see Methods). **(B)** Histogram of the first component of the PCA using peak-to-trough duration and time to repolarization to separate neurons into putative interneurons and pyramidal cells. **(C)** Proportion of non-specific PE encoding neurons identified as narrow- or broad-spiking. **(D)** Proportion of feature-specific PE encoding neurons identified as narrow- or broad-spiking or unidentified. **(E)** *Left:* Ratio of narrow to broad spiking neurons identified in the population, for non-specific and feature-specific PE encoding neurons. *Right:* High ratio of narrow/broad spiking units for feature-specific PE encoding neurons split into ACC and PFC neurons. In ACC in particular, the ratio of narrow to broad spiking units was significantly larger than expected based on the population distribution. Two black asterisks indicate $p < 0.01$ (chi-square test). **(F)-(J)** for CD/VS units. **(F)** Waveforms of all highly isolated single units recorded, identified as putative interneurons (red) or putative medium spiny neurons (MSNs, blue), or unidentified (black) (see Methods). **(G)** Histogram of the first component of the PCA using peak width and initial slope of valley decay (ISVD) to separate neurons into putative interneurons and MSNs. Inset shows the distribution of peak width to ISVD across neurons. **(H)** Proportion of non-specific PE

encoding neurons identified as putative interneurons or MSNs. **(I)** Proportion of feature-specific PE encoding neurons identified as putative interneurons or MSNs or unidentified. **(J)** Ratio of put. Interneuron/ MSN in the population, for non-specific and feature-specific PE encoding neurons.

For *monkey H* we found that, based on the distribution of narrow- and broad-spiking neurons observed in the population of PFC/ACC neurons (Figure 3.11A, B), feature-specific PE neurons contained a larger proportion of narrow-spiking neurons (ratio narrow/broad = 0.83) than would have been expected based on the distribution of neurons recorded (ratio narrow/broad in population = 0.38) (Chi-square test, p = .0036; Figure 3.11D, E left) (this ratio was qualitatively similar across different types and features of feature-specific PEs). Splitting units further into ACC and PFC neurons revealed that this effect was primarily driven by ACC (ratio narrow/broad = 1.1) (Chi-square test, p = .0041; Figure 3.11E right). Notably, neurons that encoded non-specific PE signals in PFC and ACC, were classified as narrow- and broad-spiking units in proportions as expected based on the population (Chi-square test, p = .72; Figure 3.11C, E left). Feature-specific and non-specific PE signals in CD and VS were encoded by putative interneurons and putative MSNs in proportions similar to what would be expected based on the distribution of putative interneurons and MSNs that was recorded (Chi-square test, p_{feat} = 0.30, p_{non} = 0.94; Figure 3.11G-J).

For *monkey K*, we found a similar trend of effects, that was however not statistically significant, whereby feature-specific PE units tended to be encoded by a higher ratio of putative interneurons to putative principal cells than expected based on the population in frontal (Chi-square test, p = 0.16) regions, which was not the case for non-specific PE units (Chi-square test, p = 0.79) (Figure 3.12E). Feature-specific and non-specific PE signals in CD and VS were encoded in proportions similar to what would be expected based on the distribution of putative interneurons and MSNs that was recorded (Chi-square test, p_{feat} = 0.22, p_{non} = 0.58; Figure 3.12J). It should be noted that the population ratio of putative interneurons to putative principal cells recorded in *monkey K* was smaller compared with this ratio in *monkey H*,

especially in frontal regions (*monkey H*: 0.38, *monkey K*: 0.22, Figure 3.11E, J, Figure 3.12E, J), and the resulting small population sample sizes likely contributed to the fact that these effects were not significant in *monkey K*. The previous suggests that, robustly in one animal, feature-specific PE signals in ACC were encoded more often by putative interneurons than would be expected by chance.

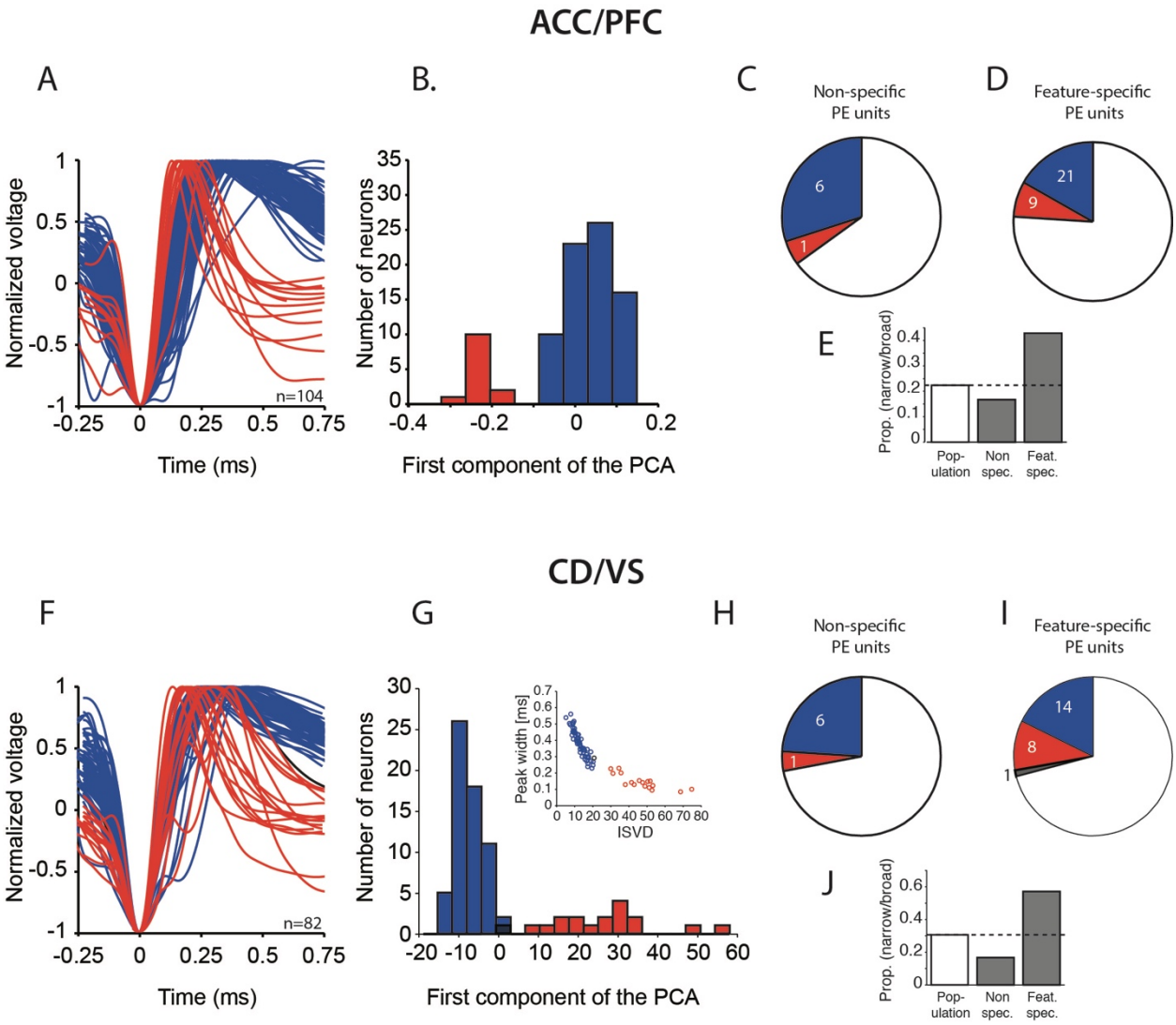

Figure 3.12 Cell-types encoding prediction errors in *monkey K*.

Conventions are equivalent to Figure 3.11.

3.5 Discussion

Using a feature-based reinforcement learning approach, we found that neurons across areas of the fronto-striatal circuit in two monkeys encode prediction error (PE) signals that are informative of those features that were selected in the preceding choice. Specifically, using a feature-based reversal learning task, we found that i) surprise signals that were uninformative about the valence of the outcome were most prevalent, encoded in all areas, and for all stimulus features, ii) positive PE signals selective for the reward relevant dimension – color – were the most prevalent positive PE signals and encoded in all areas tested, iii) color-specific negative PE signals were consistently encoded in VS, iv) PFC was the only area that encoded positive PE signals selective for all stimulus features, and v) feature-specific PE signals emerged later than non-specific PE signals in all areas.

The encoding of feature non-specific PEs and of choice-relevant information has previously been observed in all areas that were tested in this study (Kawagoe et al., 1998; Lau and Glimcher, 2007; Matsumoto et al., 2007; Gläscher et al., 2010; Asaad and Eskandar, 2011; Hayden et al., 2011; Kaping et al., 2011; Kennerley et al., 2011; Kim and Hikosaka, 2013; Wittmann et al., 2016; Asaad et al., 2017). In line with this, we observed the encoding of non-specific PEs (Figure 3.5A) and choice-specific information (Figure 3.2, S2) in all areas. However, what had not been clear so far was whether PE signals can carry sufficiently detailed information about the specific stimulus features that were chosen to potentially serve as selective learning signals. Importantly, in our task design, stimuli were characterized by three features at all times (color, location and motion direction), and all three features were relevant within a single-trial, while, however, *only* the color feature was relevant over many trials and directly linked to reward outcome. We find that indeed the vast majority of neurons that encode a PE signal do so only, or more strongly, for one of two chosen feature values, thereby carrying explicit information about the preceding feature selection (Figure 3.6, Figure 3.7), and these feature-specific PE signals were encoded more strongly than non-specific PE signals. Interestingly, we found that the most prevalent

feature-specific PE signal in all areas was unsigned, therefore did not carry information about the valence of the current outcome, and was encoded selectively for each of the features that characterized the choice stimuli, whether relevant to the reward outcome or not (Figure 3.7B, D, S3). This complements the finding that all outcome-relevant and non-relevant task variables, including color, -location and -motion direction linked to the current choice or previous choice were reliably encoded at the time of feedback, and in all areas (Figure 3.2, S2). It has previously been shown that non-relevant task variables and features are represented and tracked over multiple trials in the brain, in particular in prefrontal regions (Lauwereyns et al., 2001; Genovesio et al., 2014; Donahue and Lee, 2015). It has been argued that the encoding of non-relevant information may promote behavioral flexibility in volatile task environments by facilitating the detection of changes in this environment (Mante et al., 2013; Donahue and Lee, 2015). Surprise signals (unsigned PEs) reflect the absolute value of the prediction error, independent of whether the outcome was unexpectedly bad or unexpectedly good. Thus, feature-specific surprise signals reflect the accuracy with which a feature predicts an outcome, and therefore index the predictive uncertainty associated with that feature (Courville et al., 2006). Theories of attentional learning have suggested that the degree to which events are surprising governs learning and behavior (Kamin, 1969; Mackintosh, 1975; Pearce and Hall, 1980). Specifically, increased predictive uncertainty, associated for instance with a feature, should increase neural resources to enhance their processing and thereby govern learning about them (Pearce and Hall, 1980; Courville et al., 2006). Additionally, greater uncertainty generally promotes risk taking and exploratory choices (Doya, 2008). In light of the volatility and complexity of the task design we employed in this study, it seems possible that widely represented feature-specific surprise signals, especially in ACC (Figure S3C Appendix A), served to increase attention to all stimulus features during periods of uncertainty to promote correct behavioral adjustment. Although not tested here, it is possible that the local integration of these surprise signals differed based on the outcome-relevance of the stimulus features they represented (Donahue and Lee, 2015).

In contrast to the ubiquitously represented feature-specific surprise signals, consistently across monkeys we found that ACC, CD and VS encoded color – selective positive PE signals more often than location- or motion-selective positive PE signals (Figure 3.7B, D). This suggests that these areas selectively represented detailed information about the specific color choice, which was the reward relevant feature dimension, that preceded an unexpected rewarded outcome, and were therefore likely instrumental for successful reversal learning in this task. This is in line with a role of ACC in monitoring and guiding flexible adjustments in behavior, especially in volatile environments (Behrens et al., 2007; Alexander and Brown, 2011; Shenhav et al., 2013). ACC has also been found to encode rule information (Johnston et al., 2007; Womelsdorf et al., 2010); and activity in medial PFC (including ACC) has been shown to reflect self-applied rule switches based on color prior to the actual switch (Schuck et al., 2015), suggesting that ACC may be a driving force in adjusting rule-based behavior. In line with this, ACC has also been shown to encode value information about stimuli and actions (Matsumoto et al., 2007; Quilodran et al., 2008; Kaping et al., 2011; Klein-Flügge et al., 2016; Rudebeck et al., 2017).

VS is most commonly associated with encoding PEs in the fMRI literature (e.g. McClure et al., 2003; Hare et al., 2008; Daw et al., 2011; Niv et al., 2012), perhaps due to its strong connection with midbrain dopamine neurons (Haber and Knutson, 2010), and has often been suggested to encode stimulus values (Clarke et al., 2008; Costa et al., 2016; Rothenhoefer et al., 2017), suggesting that VS may be particularly well positioned to encoding feature-relevant PEs in this task. Neurons in CD have traditionally been associated with encoding of action values rather than stimulus values (Lau and Glimcher, 2008; Seo et al., 2012), although the encoding of stimulus values has also been shown (Kim and Hikosaka, 2013; Vo et al., 2014; Hikosaka et al., 2017). We find evidence here that when color is the reward-relevant feature, neurons in CD preferentially encode PE signals that are informative about the preceding color choice, suggesting that CD neurons are similarly involved in encoding relevant value information in this domain.

In contrast to ACC, CD and VS, we find that PFC neurons consistently in both animals encoded positive PE signals selective for each of the three features that characterized the stimuli - reward-relevant and non-

relevant (Figure 3.7B, D). This is in line with the finding that the encoding of non-relevant task information has predominantly been reported in PFC neurons (Lauwereyns et al., 2001; Genovesio et al., 2014; Donahue and Lee, 2015), and PFC has previously also been shown to encode PEs selective for the preceding stimulus choice (Asaad et al., 2017). We add to this by showing that PFC neurons encode positive PEs selective for the preceding feature choices independent of their relevance to reward-guided behavior, potentially to promote continuous behavioral flexibility in volatile environments (Donahue and Lee, 2015).

Feature-specific negative PEs were almost exclusively encoded for the reward-relevant color dimension (Figure 3.7B, D), potentially assigning them particular importance in driving correct behavioral adjustments in this task. These color-specific negative PEs were consistently found in the ventral striatum of both animals (Figure 3.7B, D, S3), suggesting that the VS may be instrumental in signaling the specific feature choices that led to unexpectedly negative and positive outcomes.

We generally observed negative PE encoding, whether feature- or non-specific, less often than positive PE and surprise encoding (Figure 3.5A, Figure 3.7A-D). Some previous studies that have used reversal learning tasks with multiple possible targets found that animals learned poorly from incorrect choices and indeed returned to chance level performance following an incorrect outcome, even though that information could have been used towards inferring the target identity (Fusi et al., 2007; Asaad and Eskandar, 2011). It has been suggested that this type of behavior may reflect behavioral adaptation to volatile task environments (Asaad and Eskandar, 2011). It may be possible that this bias in behavior is reflected in less neurons encoding negative PEs. It should be noted, however, that error trials were less commonly observed than correct trials in this task, and therefore a bias in statistical power may have additionally contibuted to the fact that fewer negative PE signals were observed.

We found that across areas feature-specific PE signals generally emerged later than non-specific PE signals and reward-encoding per se. This could potentially suggest that feature-specific PE units integrate information about reward expectations, violation of expectations and preceding feature choice from potentially different sources (Tian et al., 2016) leading to an increase in processing time.

With regards to the latency with which feature-specific PE signals emerged in the different areas, we could not identify a single time course of encoding (Figure 3.8, 3.9, S7, S8). Rather, we find that order of encoding differed quite drastically with the type of signal that was encoded as well as with the individual animal. Color-specific positive PEs for instance were encoded soonest in ACC and VS in *monkey H* and soonest in CD and ACC in *monkey K*, while feature-specific unsigned PEs were encoded soonest in CD and ACC in *monkey H,* and soonest in PFC in *monkey K* (Figure 3.8, 3.9). Smaller sample sizes in *monkey K* may have contributed to the difference in latencies we observe between the two animals. It is however equally, if not more, likely that individual differences contributed to these differing results, since we observed that the encoding of reward per se, which occurred in a large fraction of cells, and even though it was independent of the latency of PE encoding (Figure 3.10), differed fundamentally between the two animals – in *monkey H,* the two frontal regions encoded reward soonest, in *monkey K*, the two striatal regions encoded reward soonest (Figure 3.10). The previous could suggest the possibility that the two animals at times relied on different but overlapping approaches to learning in this task (Gläscher et al., 2010; Daw et al., 2011; Diuk et al., 2013; Guo et al., 2016). In tasks that employ two stages of learning (e.g. hierarchical), PE signaling can be dissociated based on the stage of learning in striatal and frontal regions (Daw et al., 2011; Diuk et al., 2013). PE signals in the striatum for instance have been shown to depend on whether goal-directed reward associations are learned versus whether stimulus associations are learned in a task in which both had to be learned independently to achieve reward (Guo et al., 2016). Even though we cannot explicitly dissociate these two types of learning in our task, a response in our task was contingent on the motion direction of a stimulus, while reward was contingent on the color of the stimulus. Although speculative, it might be possible that two associations operated in parallel

(color-reward and stimulus-response) and that a given animal in this task may have relied at times more on one versus the other.

The heavy interconnecting network of subregions of the frontal cortex and striatum, as well as the direct projections from midbrain dopamine neurons to striatal and frontal regions (Haber and Knutson, 2011), make it potentially challenging to identify one specific area that leads the others in encoding feature-specific PE signals. Few studies in the past have examined PE signals at the neural level in more than one area (e.g. Sul et al., 2010; Asaad and Eskandar, 2011; Kennerley et al., 2011), and even fewer report on differences in the latency of these signals, although Asaad and Eskandar (2011) report that PE signals emerged with similar latencies in caudate and dorsolateral PFC. The previous suggests that all areas we tested here are involved in encoding PE signals that carry relevant feature choice information, but differences in dominance of one area over another may depend on differences in the type of signal, such as valence-content, and could depend on individual approaches to learning.

We found, robustly in one animal and as a trend in the second animal, that in frontal regions and particularly in ACC, feature-specific PE signals compared with non-specific PE signals, were encoded predominantly by putative interneurons, while in all other regions the distribution of putative interneurons to putative principal cells that encoded feature-specific PEs was similar to what would be expected based on the ratio of observing these neuron types (Figure 3.11, 3.12). Since this effect is area-specific, it is unlikely explainable based on higher firing rates of putative interneurons, in which case we would expect the same results for PFC and striatal regions. Recent observations suggest that ACC, compared with PFC, has a more extensive inhibitory circuitry, with a greater frequency and size of inhibitory postsynaptic currents and a greater density and diversity of inhibitory synapses (Medalla et al., 2017). The greater involvement of ACC inhibitory neurons in the encoding of feature-specific PE signals could therefore indicate the greater engagement of control signals (Alexander and Brown, 2011; Shenhav et al., 2013) elicited by these performance- and choice-informative signals to support the successful maintenance or

adjustment of current behavioral strategies (Buckley et al., 2009), that are relayed to other frontal and subcortical areas (Medalla and Barbas, 2009, 2010; Haber and Knutson, 2010).

The multiplexing of information about outcome and preceding choice relates to the problem of credit assignment, which describes the problem of correctly assigning credit for a given outcome to the source of that outcome in the preceding choice. The fact that we widely observe signals that are informative about how accurate the current outcome was predicted as well as about the specific feature that was selected in the preceding choice, suggests that these signals could serve as a source for solving the credit assignment problem. Our data suggests that neurons in fronto-cingulate and striatal regions may be of particular importance to this process, which in turn suggests that this functionality cannot solely be attributed to regions of the frontal cortex, particularly OFC, which has prominently been shown to be essential for solving the credit assignment problem (Walton et al., 2011; Chau et al., 2015; Jocham et al., 2016; Noonan et al., 2017).

The preceding evidence suggests that multiplexing of relevant task information at the neuron level is remarkably common and extends to the multiplexing of information about unexpected outcomes and preceding stimulus feature choices, which may aid the problem of credit assignment. Indeed, most neurons that encoded a prediction error also encoded information about a previously selected feature, suggesting that these inherently more informative signals are commonly encoded in many brain regions and may constitute the preferred type of signal encoding. We speculate that based on the reward-relevance of the feature-specific PE signal, they may lead to differential changes in attention and their precision or reliability may be weighted in a way that allows identifying of those synapses that require updating during learning (Friston, 2005; Roelfsema et al., 2010; den Ouden et al., 2012). This also fits

with the recent notion that PE signals seem to be encoded brain-wide and in an almost redundant fashion (Den Ouden et al., 2012; Tian et al., 2016).

Future studies should investigate whether different areas of the frontal and striatal circuit are differentially involved in encoding feature-specific PEs when the reward-relevant feature switches between e.g. color, location, and motion direction. This may reveal dissociable contributions of e.g. caudate and ACC neurons in encoding action-specific PEs when actions are reward-relevant (Matsumoto et al., 2007; Stalnaker et al., 2012) and PFC in location-specific PEs when spatial locations are reward-relevant (Lara and Wallis, 2014).

Overall, we find that prediction error signals that are uniquely informative about the preceding feature choice are remarkably commonly represented across the fronto-striatal circuit. Valence uninformative surprise signals may cause an increase in attention to all stimulus features during periods of uncertainty, while color-specific positive and negative prediction error signals in ACC, caudate and ventral striatum may selectively help assign credit and identify those synapses that require updating to allow flexible adjustments in goal-directed behavior.

3.6 References

Akaishi R, Kolling N, Brown JW, Rushworth M (2016) Neural Mechanisms of Credit Assignment in a Multicue Environment. J Neurosci 36:1096–1112.

Akatsuka K, Wasaka T, Nakata H, Kida T, Kakigi R (2007) The effect of stimulus probability on the somatosensory mismatch field. Exp Brain Res 181:607–614.

Alexander WH, Brown JW (2011) Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14:1338–1344.

Ardid S, Vinck M, Kaping D, Marquez S, Everling S, Womelsdorf T (2015) Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. J Neurosci 35:2975–2991.

Asaad WF, Eskandar EN (2011) Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J Neurosci 31:17772–17787.

Asaad WF, Lauro PM, Perge JA, Eskandar EN (2017) Prefrontal neurons encode a solution to the credit assignment problem. J Neurosci 37:3311–3316.

Bach DR, Symmonds M, Barnes G, Dolan RJ (2017) Whole-brain neural dynamics of probabilistic reward prediction. J Neurosci 37:2943–16.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Behrens TE, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Berke JD (2008) Uncoordinated firing rate changes of striatal fast-spiking interneurons during behavioural task performance. J Neurosci 28:10075–10080.

Berke JD, Okatan M, Skurski J, Eichenbaum HB (2004) Oscillatory entrainment of striatal neurons in freely moving rats. Neuron 43:883–896.

Bestmann S, Harrison LM, Blankenburg F, Mars RB, Haggard P, Friston KJ, Rothwell JC (2008) Influence of Uncertainty and Surprise on Human Corticospinal Excitability during Preparation for Action. Curr Biol 18:775–780.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PGF, Kwok SC, Phillips A, Tanaka K (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science 325:52–58.

Cai X, Padoa-Schioppa C (2014) Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. Neuron 81:1140–1151.

Chau BKH, Sallet J, Papageorgiou GK, Noonan MAP, Bell AH, Walton ME, Rushworth MFS (2015) Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. Neuron 87:1106–1118.

Clarke HF, Robbins TW, Roberts AC (2008) Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. J Neurosci 28:10972–10982.

Corlett PR, Aitken MRF, Dickinson A, Shanks DR, Honey GD, Honey RAE, Robbins TW, Bullmore ET, Fletcher PC (2004) Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. Neuron 44:877–888.

Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB (2016) Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. Neuron 92:505–517.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends Cogn Sci 10:294–300.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.

Dayan P, Niv Y (2008) Reinforcement learning: The Good, The Bad and The Ugly. Curr Opin Neurobiol 18:185–196.

den Ouden HEM, Kok P, de Lange FP (2012) How prediction errors shape perception, attention, and motivation. Front Psychol 3:1–12.

Den Ouden HEM, Kok P, de Lange FP (2012) How prediction errors shape perception, attention, and motivation. Front Psychol 3:1–12.

Diuk C, Tsai K, Wallis J, Botvinick M, Niv Y (2013) Hierarchical Learning Induces Two Simultaneous, But Separable, Prediction Errors in Human Basal Ganglia. J Neurosci 33:5797–5805.

Donahue CH, Lee D (2015) Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. Nat Neurosci 18:1–9.

Doya K (2008) Modulators of decision making. Nat Neurosci 11:410–416.

Dunn OJ, Clark VA (1987) Applied Statistics: Analysi of Variance and Regression. :175–178.

Everling S, Tinsley CJ, Gaffan D, Duncan J (2002) Filtering of neural signals by focused attention in the monkey prefrontal cortex. Nat Neurosci 5:671–676.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science (80-) 299:1898–1902.

Friston KJ (2005) A theory of cortical responses. Philos Trans R Soc B 360:815–836.

Fusi S, Asaad WF, Miller EK, Wang XJ (2007) A Neural Circuit Model of Flexible Sensorimotor Mapping: Learning and Forgetting on Multiple Timescales. Neuron 54:319–333.

Genovesio A, Tsujimoto S, Navarra G, Falcone R, Wise SP (2014) Autonomous Encoding of Irrelevant Goals and Outcomes by Prefrontal Cortex Neurons. J Neurosci 34:1970–1978.

Glantz S, Slinker B (2001) Primer of Applied Regression and Analysis of Variance. :25–28.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.

Guo R, Bohmer W, Hebart M, Chien S, Sommer T, Obermayer K, Gläscher J (2016) Interaction of Instrumental and Goal-Directed Learning Modulates Prediction Error Representations in the Ventral Striatum. J Neurosci 36:12650–12660.

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci 28:5623–5630.

Hassani SA, Oemisch M, Balcarras M, Westendorff S, Ardid S, van der Meer MA, Tiesinga P, Womelsdorf T (2017) A computational psychiatry approach identifies how alpha-2A noradrenergic agonist Guanfacine affects feature-based reinforcement learning in the macaque. Sci Rep 7:40606.

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. J Neurosci 31:4178–4187.

Hikosaka O, Ghazizadeh A, Griggs W, Amita H (2017) Parallel basal ganglia circuits for decision making. J Neural Transm:1–15.

Jocham G, Brodersen KH, Constantinescu AO, Kahn MC, Ianni AM, Walton ME, Rushworth MFS, Behrens TEJ (2016) Reward-Guided Learning with and without Causal Attribution. Neuron 90:177–190.

Johnston K, Levin HM, Koval MJ, Everling S (2007) Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron 53:453–462.

Kamin LJ (1969) Predictability, surprise, attention, and conditioning. In: Punishment and Aversive Behavior (Church R, Campbell B, eds), pp 279–296. New York: Appleton-Century-Crofts.

Kaping D, Vinck M, Hutchison RM, Everling S, Womelsdorf T (2011) Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. PLoS Biol 9:e1001224.

Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. 1:411–416.

Kennerley SW, Behrens TEJ, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581–1589.

Kim H, Hikosaka O (2013) Distinct Basal Ganglia Circuits Controlling Behaviors Guided by Flexible and Stable Values. Neuron 79:1001–1010.

Klein-Flügge MC, Kennerley SW, Friston K, Bestmann S (2016) Neural signatures of value comparison in human cingulate cortex during decisions requiring an effort-reward trade-off. J Neurosci 36:10002–10015.

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205.

Lansink CS, Goltstein PM, Lankelma J V, Pennartz CMA (2010) Fast-spiking interneurons of the rat ventral striatum: temporal coordination of activity with principal cells and responsiveness to reward. Eur J Neurosci 32:494–508.

Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. J Neurosci 27:14502–14514.

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463.

Lauwereyns J, Sakagami M, Tsutsui K, Kobayashi S, Koizumi M, Hikosaka O (2001) Responses to task-irrelevant visual features by primate prefrontal neurons. J Neurophysiol 86:2001–2010.

Lebedev M a, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y (2017) Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron 93:451–463.

Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev 82:276–298.

Mante V, Sussillo D, Shenoy K V, Newsome WT (2013) Context-dependent computation by recurrent

dynamics in prefrontal cortex. Nature 503:78–84.

Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647–656.

McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346.

Medalla M, Barbas H (2009) Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. Neuron 61:609–620.

Medalla M, Barbas H (2010) Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. J Neurosci 30:16068–16081.

Medalla M, Gilman JP, Wang J-Y, Luebke JI (2017) Strength and diversity of inhibitory signaling differentiates primate anterior cingulate from lateral prefrontal cortex. J Neurosci 37:3757–16.

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, Wilson RC (2015) Reinforcement learning in multidimensional environments relies on attention mechanisms. J Neurosci 35:8145–8157.

Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci 32:551–562.

Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J Neurosci 30:10692–10702.

Noonan MP, Chau B, Rushworth MF, Fellows LK (2017) Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision making in humans. J Neurosci 37:7023–7035.

Oemisch M, Westendorff S, Everling S, Womelsdorf T (2015) Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. J Neurosci 35:13076–13089.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223–226.

Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat Neurosci 11:95–102.

Pearce JM, Hall G (1980) A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli. Psychol Rev 87:532–552.

Quilodran R, Rothé M, Procyk E (2008) Behavioral Shifts and Action Valuation in the Anterior Cingulate Cortex. Neuron 57:314–325.

Roelfsema PR, van Ooyen A, Watanabe T (2010) Perceptual learning rules based on reinforcers and attention. Trends Cogn Sci 14:64–71.

Rothenhoefer KM, Costa VD, Bartolo R, Vicario-Feliciano R, Murray EA, Averbeck BB (2017) Effects of ventral striatum lesions on stimulus versus action based reinforcement learning. J Neurosci 37:0631–17.

Rudebeck PH, Ripple JA, Mitz AR, Averbeck BB, Murray EA (2017) Amygdala Contributions to Stimulus–Reward Encoding in the Macaque Medial and Orbital Frontal Cortex during Learning. J Neurosci 37:2186–2202.

Rushworth MFS, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. Trends Cogn Sci 8:410–417.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.

Schuck NW, Gaschler R, Wenke D, Heinzle J, Frensch PA, Haynes JD, Reverberi C (2015) Medial Prefrontal Cortex Predicts Internally Driven Strategy Shifts. Neuron 86:1–10.

Schultz W (1998) Predictive Reward Signal of Dopamine Neurons. J Neurophysiol 80:1–27.

Schultz W, Apicella P, Ljungberg T (1993) Responses of Monkey Dopamine Neurons to Reward and Conditioned Stimuli during Successive Steps of Learning a Delayed Response Task. J Neurosci 13:900–913.

Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci 27:8366–8377.

Seo M, Lee E, Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. Neuron 74:947–960.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. Science (80-) 348:1352–1355.

Smith AC, Brown EN (2003) Estimating a State-Space Model from Point Process Observations. Neural Comput 15:965–991.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki W a, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. J Neurosci 24:447–461.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2012) Reward prediction error signaling in posterior dorsomedial striatum is action-specific. J Neurosci 32:10296–10305.

Sul JH, Kim H, Huh N, Lee D, Jung MW (2010) Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron 66:449–460.

Summerfield C, Koechlin E (2008) A Neural Representation of Prior Information during Perceptual Inference. Neuron 59:336–347.

Sutton RS, Barto A (1998) Reinforcement Learning. Cambridge, MA: MIT Press.

Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, Watabe-Uchida M (2016) Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. Neuron:1374–1389.

Vo K, Rutledge RB, Chatterjee A, Kable JW (2014) Dorsal striatum is necessary for stimulus-value but not action-value learning in humans. Brain 137:3129–3135.

Walton ME, Behrens TEJ, Noonan MP, Rushworth MFS (2011) Giving credit where credit is due : orbitofrontal cortex and valuation in an uncertain world. Ann N Y Acad Sci 1239:14–24.

Westendorff S, Kaping D, Everling S, Womelsdorf T (2016) Prefrontal and anterior cingulate cortex neurons encode attentional targets even when they do not apparently bias behavior. J Neurophysiol 116:796–811.

Wilson RC, Niv Y (2012) Inferring Relevance in a Changing World. Front Hum Neurosci 5:1–14.

Wittmann MK, Kolling N, Akaishi R, Chau BKH, Brown JW, Nelissen N, Rushworth MFS (2016) Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nat Commun 7:12327.

Womelsdorf T, Johnston K, Vinck M, Everling S (2010) Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. Proc Natl Acad Sci 107:5248–5253.

Chapter 4

Changes of attention during value-based reversal learning are tracked by N2pc and feedback-related negativity

Mariann Oemisch[1], Marcus Watson[1], Thilo Womelsdorf[1,2], Anna Schubö[3]

[1]Department of Biology, Centre for Vision Research, York University, Toronto, Ontario M6J 1P3, Canada; [2]Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240, USA; [3]Department of Psychology, Philipps-Universität Marburg, 35032 Marburg, Germany

in press at Frontiers in Human Neuroscience

4.1 Summary

Previously learned reward values can have a pronounced impact, behaviorally and neurophysiologically, on the allocation of selective attention. All else constant, stimuli previously associated with a high value gain stronger attentional prioritization than stimuli previously associated with a low value. The N2pc, an ERP component indicative of attentional target selection, has been shown to reflect aspects of this prioritization, by changes of mean amplitudes closely corresponding to selective enhancement of high value target processing and suppression of high value distractor processing. What has remained unclear so far is whether the N2pc also reflects the flexible and repeated behavioral adjustments needed in a volatile task environment, in which the values of stimuli are reversed often and unannounced. Using a value-based reversal learning task, we found evidence that the N2pc amplitude flexibly and reversibly tracks value-based choices during the learning of reward associated stimulus colors. Specifically, successful learning of current value-contingencies was associated with reduced N2pc amplitudes, and this effect was more apparent for distractor processing, compared with target processing. In addition, following a value reversal the FRN, an ERP component that reflects feedback processing, was amplified and co-occurred with increased N2pc amplitudes in trials following low-value feedback. Importantly, participants that showed the greatest adjustment in N2pc amplitudes based on feedback were also the most efficient learners. These results allow further insight into how changes in attentional prioritization in an uncertain and volatile environment support flexible adjustments of behavior.

4.2 Introduction

Visual selective attention allows the prioritization of task-relevant over irrelevant stimuli in the visual field. Traditionally, selective attention has been dissociated into goal-directed 'top-down' driven selective attention and salience-driven 'bottom-up' selective attention (e.g. Posner and Petersen, 1990; Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002). However, in recent years it has become evident that this

dichotomy does not suffice to explain all instances in which a stimulus, or a set of stimuli, become the target of attentional priority (Awh et al., 2012; Anderson, 2013; Womelsdorf and Everling, 2015). A third source of attentional control, referred to as 'experience-driven', includes an individual's recent selection-history and reward learning (Della Libera and Chelazzi, 2006; Anderson et al., 2011a; Awh et al., 2012; Irons and Leber, 2016). In particular, previously learned reward value has been shown to be a strong modulator of attentional prioritization (e.g. Della Libera and Chelazzi, 2009; Krebs et al., 2010; Anderson et al., 2011b, 2013, 2014; Della Libera et al., 2011; Hickey et al., 2011, 2015; Sali et al., 2014; Bucker and Theeuwes, 2016). For example, non-salient and task-irrelevant distractors that have previously been associated with reward can capture attention involuntarily and cause slower reaction times in classical visual search tasks, and this is modulated by reward level, such that the higher the previously-associated reward, the greater the subsequent capture (e.g. Anderson et al., 2011b, 2013; Munneke et al., 2015). Neurophysiological evidence for this increased attentional capture by high-valued stimuli has been found in the mean amplitude of the N2pc (e.g. Kiss et al., 2009; Feldmann-Wüstefeld et al., 2015, 2016; Sawaki et al., 2015). The N2pc is an EEG component thought to reflect attentional target selection processes (Luck and Hillyard, 1994a; Woodman and Luck, 2003; Eimer, 2014; Eimer and Grubert, 2014), likely generated in intermediate and high levels of the ventral visual processing pathway (Hopf et al., 2000; Hopf, 2006). It onsets earlier and is more pronounced when search targets are associated with a higher value (Kiss et al., 2009), and the presence of higher value distractors causes a decrease in target enhancement and an increase in distractor suppression during visual search (Feldmann-Wüstefeld et al., 2016). Sawaki and colleagues (2015) found that prior to visual search, a high value cue elicited stronger distractor suppression than a low value cue, and thereafter a smaller N2pc component was observed during the visual search. The authors argue that increased suppression to a high value but non-informative (to target selection) cue may have allowed better performance in the following visual search, which was supported by decreased reaction times as well as decreased alpha oscillation levels prior to the visual search that indicated heightened visual readiness (Sawaki et al., 2015).

We have thus already gained substantial insight into the behavior and neurophysiological processes that underlie the selective processing of high- or low-valued target and distractor signals. However, the distinction between targets and distractors is often much less clear outside the classic experimental environment. Real life is substantially more volatile, and therefore the stimuli that surround us must continuously be reevaluated with regards to their relevance to our current goals. A critical goal of attentional prioritization is likely reward maximization and loss minimization (e.g. Navalpakkam et al., 2010), meaning that in a dynamic world we have to continuously learn and update our choice criteria with regards to the stimuli we attend.

We do not yet know how flexibly attentional target selection, as tracked by the N2pc component, can change in response to changes in reward values. In this study, we therefore employed a value-based reversal learning task in which stimulus reward values changed repeatedly and without warning. Specifically, participants were asked to attend to and choose one of two target stimuli that differed in color and likelihood of leading to a high reward outcome. This color-value association changed often and unannounced, such that the previously high value stimulus became the low value stimulus and vice versa. Participants therefore had to continuously re-evaluate, based on trial and error, whether they were choosing the currently high valued stimulus. This allowed us to assess learning-related changes in behavior, and to compare neural processing when subjects were in the process of learning the current value contingency, with processing when they had already successfully learned the current value contingency. Using EEG recordings, we examined learning- and choice-related differences to the N2pc. Since participants used trial-by-trial feedback to evaluate choices, we also examined feedback-related differences to the N2pc and learning-related differences to frontal feedback related negativity (FRN), an EEG component observed over fronto-central sites that is usually measured as a voltage difference following positive versus negative feedback (Cohen et al., 2011). The FRN has previously been shown to change during reversal learning and has been suggested to encode prediction error signals and behavioral adjustments (e.g. Chase et al., 2010; Cohen and Ranganath, 2007; Walsh and Anderson, 2011).

Neural processing of the valued stimuli was isolated by always placing one valued stimulus on the vertical midline, thereby attributing the lateralized EEG activity to the second, lateralized stimulus (e.g. Hickey et al., 2009; Feldmann-Wüstefeld et al., 2016). Importantly, our task design did not have a fixed dissociation into 'target' and 'distractor', since either of the two stimuli could be selected for response and the identity of the high and low value stimuli changed frequently. Instead, we differentiated processing of the selected (target) and the non-selected (distractor) stimulus on a trial-by-trial basis dependent on participants' choice performance. We expected to observe learning-dependent changes in attentional prioritization reflected in N2pc amplitudes, and in feedback-processing reflected in FRN amplitudes (Figure 4.1). When performing similar tasks (e.g. Cools et al., 2002; Chase et al., 2010) participants generally show a low probability of choosing the newly high-valued stimulus in the trials immediately following a value reversal and therefore *during* learning of the current value contingency, and a high probability of choosing the currently high valued stimulus *after* successful learning (Figure 4.1A). We hypothesized that changes in N2pc and FRN amplitudes would parallel these changes in learning behavior. Specifically, we hypothesized that with successful learning of the current value contingency, the N2pc elicited by the stimulus selected for response (and therefore presumably actively attended), should increase, potentially reflecting more precise attentional selection of the current target stimulus (Figure 4.1B left). We furthermore hypothesized that the N2pc elicited by the stimulus that was *not* selected for response (therefore presumably not actively attended) should generally be smaller than that of the stimulus that was selected for response, and it should further decrease with successful learning, potentially reflecting more successful avoidance of attentional capture by the current distractor stimulus (Figure 4.1B right). Alternatively, it is possible that relatively fast learning during frequent value reversals is not reflected in changes in early attentional stimulus selection as measured with the N2pc, or that attentional processing of only the current target or only the current distractor is affected. Finally, we expected feedback processing reflected in FRN amplitudes to be greater *during* learning of the current value contingency compared with *after* learning, potentially reflecting the greater propensity to keep track

of accumulated feedback when behavior needs to potentially be adjusted following a value reversal (Figure 4.1C, Chase et al., 2010).

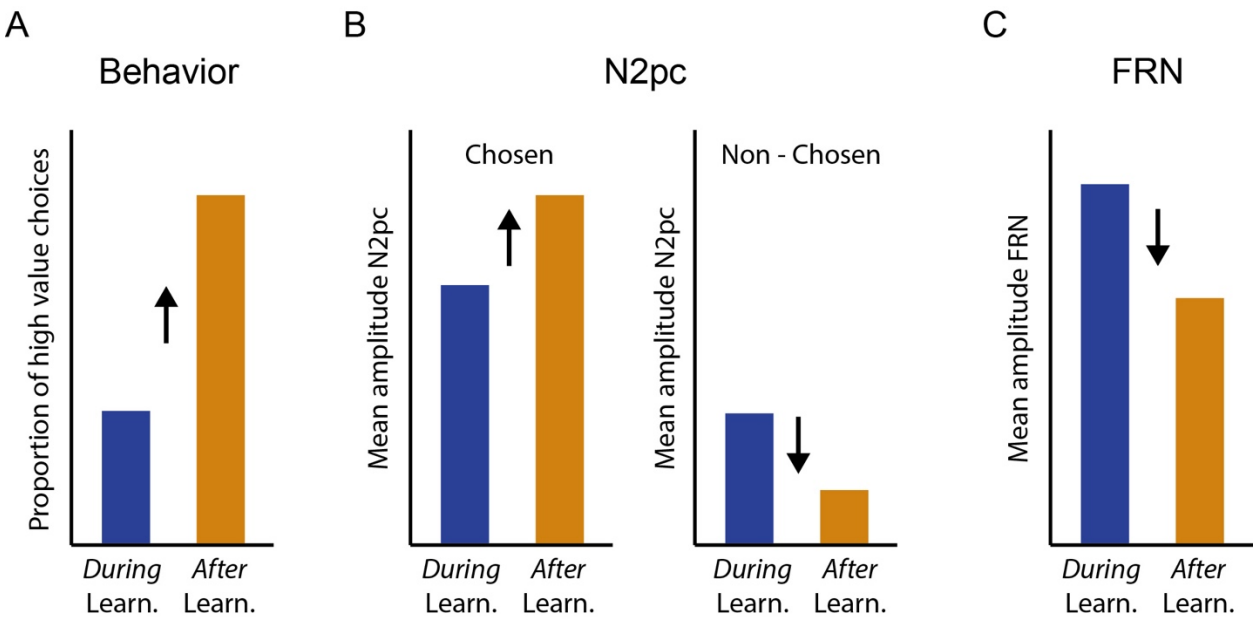

Figure 4.1 Hypotheses for learning-related changes in behavior, N2pc, and FRN amplitudes.

(A) Successful learning (*after* learning) in our value based reversal learning task is reflected in an increased probability of choosing the currently high valued target. **(B)-(C)** We expected N2pc and FRN amplitudes to change in parallel with learning behavior. We hypothesized that N2pc amplitudes for the stimulus that was chosen for response, and therefore presumably actively attended, would increase with learning, to potentially reflect more accurate attentional target selection with successful learning (**B** left). In contrast, we hypothesized that N2pc amplitudes for the stimulus that was *not* chosen for response, and therefore presumably *not* actively attended, would be substantially smaller than that for the stimulus chosen for response, and would further decrease with learning, to potentially reflect more successful avoidance of attentional capture by the distracting stimulus (**B** right). **(C)** We hypothesized that FRN amplitudes would be greater *during* learning of the current value contingency compared with *after* learning, potentially reflecting the greater propensity to actively assess feedback during periods that may require behavioral adjustment.

4.3 Methods

4.3.1 Participants

Twenty-six students of the Philipps-University Marburg participated in the experiment for payment (8€/h) or course credit. Contingent on performance, participants could collect an additional monetary bonus of up to 6€. Three participants were rejected from analysis because too many trials (> 40%) were lost due to EEG artifacts and non-learning (see **Data analysis**). Two participants had to be rejected due to technical issues. Analyses are shown for the remaining twenty-one participants (15 female, 6 male, mean age = 21.4). Three out of those twenty-one participants were left-handed. All participants were naïve to the purpose of the experiment and had normal or corrected-to-normal visual acuity and normal color vision. Visual acuity and color vision were tested with an OCULUS Binoptometer 3 (OCULUS Optikgeräte GmbH, Wetzlar, Germany). This study was carried out in accordance with the recommendations of the Ethics Committee of the Department of Psychology at the Philipps University Marburg with written informed consent from all participants, in accordance with the Declaration of Helsinki.

4.3.2 Apparatus and stimuli

Participants were seated in a comfortable chair in a dimly lit and electrically shielded room, facing a monitor placed at a distance of approximately 100 cm from their eyes. Stimuli were presented on a 22” screen (1680 x 1050 px) using Unity3D 5.3.5 (Unity Technologies, San Francisco, CA). The display showed eight stimuli (diameter of 2.7°) arranged equidistantly on an imaginary circle with an eccentricity of 5.5° of visual angle (Figure 4.2A). All stimuli were presented against a grey background. Six of the stimuli were dark grey circles (RGB 24, 24, 24); the remaining two stimuli were one of two different target colors. Half the participants were presented with one pink (RGB 237, 83, 255) and one green (RGB

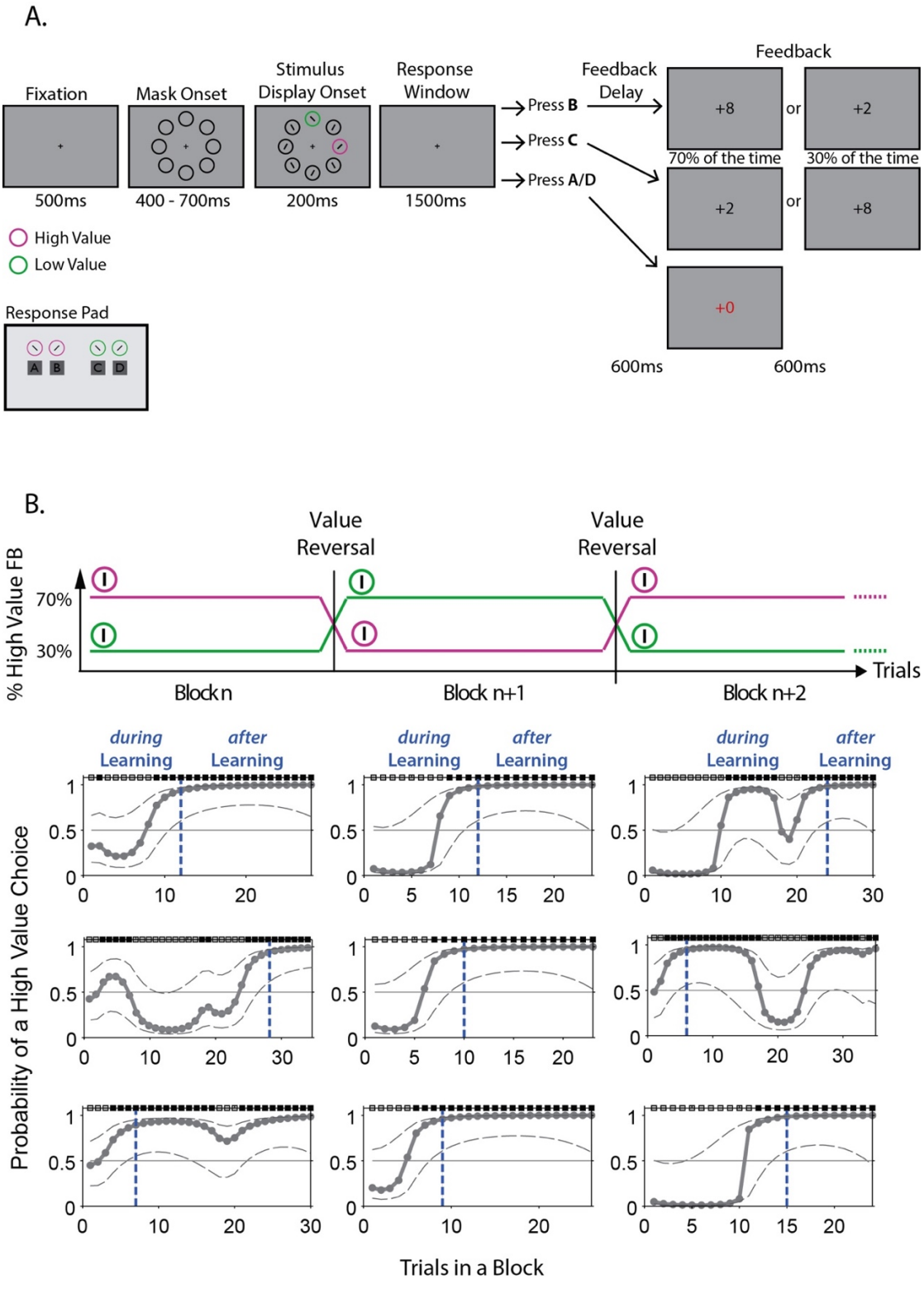

Figure 4.2 Value-based reversal learning task and example block performance computed with EM algorithm.

(A) The task started with a central fixation cross, followed by the appearance of eight black circles. For the target display, lines with different orientation appeared in all circles and two circles changed color to pink and green (or blue and orange). Participants had 1500ms to report the line orientation (+45° or -45° tilt) inside one of the two colored stimuli. Selecting the high value stimulus would lead to a high value feedback (+8) in 70% of the trials and to a low value feedback (+2) in 30% of the trials. This was reversed for the low value stimulus. Feedback was presented for 600ms. If an incorrect color-line orientation combination was reported, a +0 was shown in red font for 1200ms. After 1000-1300ms, a new trial was initiated. The response pad used by participants is illustrated below the task **(B)** Top: Schematic illustration of the value reversals applied to the colored stimuli in consecutive blocks. Bottom: Displayed is the probability of a high value choice across trials for 9 example blocks performed by one representative participant. The probability of a high value choice was computed using an Expectation Maximization (EM) algorithm (see Methods and Smith et al., 2004). Dotted grey lines represent 95% confidence intervals. Dotted blue lines indicate the trial at which learning has occurred (T_{Learn}) according to the ideal observer confidence interval. Trials are split into *during* learning and *after* learning trials according to T_{Learn}. Black and grey boxes above plots indicate high value and low value choice trials, respectively.

29, 181, 13) circle, half with one orange (RGB 217, 148, 14) and one blue (RGB 44, 168, 255) circle. All four possible stimulus colors were approximately isoluminant with the grey background (45.8 - 55.65 cd/m2, luminance background: 51.12cd/m2). Each stimulus contained a black line. Lines inside black stimuli were tilted 30° to the left or the right, alternating around the circle. Lines inside colored target stimuli were always tilted in opposite directions, by 45° to the left or 45° to the right. The two colored stimuli were always separated by one dark grey stimulus, such that one stimulus was always presented on the vertical midline either below or above the fixation cross, while the other was presented laterally to the left or right of the fixation cross. This experimental design was chosen because it allows isolating the processing related to the color stimulus presented laterally from processing of the color stimulus presented vertically. Traditionally, this design is used to isolate target-related from distractor-related processing (e.g. Hickey et al., 2009). However, we do not have a pre-defined target and distractor, rather the same color stimulus changes roles several times throughout the experiment and participants are free to decide which color stimulus they select in a trial (Irons and Leber, 2016) (see **Procedure** below). Thus, we are interested in how observers process stimuli which are associated with different reward values that change roles throughout the experiment.

Four response buttons were arranged on an Ergodex DX1 response pad, such that participants could comfortably place the middle and index fingers of both hands on the buttons. Each participant was

randomly assigned to respond to a given stimulus color with the left or right hand, e.g. participants with green and pink stimuli were randomly assigned to respond to pink stimuli with their right hand and green stimuli with their left hand, or vice versa. In either case, the left-most finger of each hand (middle finger left hand and index finger right hand) was used to respond to a 45° tilt to the left (assuming a vertical line) and the right-most finger of each hand (index finger left hand and middle finger right hand) was used to respond to a 45° tilt to the right.

4.3.3 Procedure

The task is illustrated in Figure 4.2. Participants were instructed to keep their eyes on the center of the screen throughout a trial. Each trial started with the appearance of a central fixation cross for 500ms, followed by eight dark grey circles without lines, presented for 400 – 700ms. Two circles, one on the vertical midline and one on the horizontal midline, then changed colors, and the lines appeared in all stimuli. Since lines inside colored stimuli were always tilted in opposite directions and participants were free to respond to either color stimulus, two 'correct' (high and low value) choices were possible in any given trial (e.g. pink – rightward orientation and green – leftward orientation; trial example from Figure 4.2). An incorrect response was recorded when a color/line-orientation pair was reported that was not presented in the display (e.g. pink – leftward orientation and green – rightward orientation). This stimulus display was presented for 200ms, after which participants had 1500ms to respond, followed by a delay of 600ms after which feedback was shown for another 600ms. Feedback had three possible values: a high-value '+8', a low-value '+2', or a '+0' for incorrect responses, the latter shown in red font for 1200ms. If no response was made within 1500ms, participants were asked to respond faster in the next trial via a written visual display. The inter-trial interval was 1000-1300ms.

Each stimulus display always contained the two colored stimuli (e.g. pink and green) and participants freely chose to report the line orientation of either of the two stimuli. At a given time, one color stimulus (e.g. pink) was associated with a 70% probability (high value) of leading to the outcome '+8' and a 30%

probability of leading to the outcome '+2'. The second color stimulus (e.g. green) was simultaneously associated with a 30% probability (low value) of leading to an outcome of '+8' and a 70% probability of leading to an outcome of '+2'. Across trials of a block, the color-outcome probability association remained constant for 25 to a maximum of 50 trials (e.g. color 1 high valued). After trial 25, a running average of 80% high value choices over the last 12 trials triggered a block change (e.g. now color 2 high valued), and if this did not occur by trial 50, the block change happened automatically. The reversal was unannounced, requiring the participant to use performance feedback to detect reversals.

Participants were instructed to collect as many points as possible and to respond as fast as possible without jeopardizing response accuracy. They were explicitly informed of the 70%-30% reward outcome distribution and understood that they should optimally always try to choose the color stimulus with the 70% high value outcome probability. Participants were also informed that the color-value associations would change within the experiment. Participants performed a total of 1200 trials, where stimulus positions and target line orientations were pseudo-randomly chosen on each trial. The color that was first associated with a high value was randomly chosen in each experimental session. Each experimental session (1200 trials) lasted approximately 100 minutes including a 10-minute break after 50-60 minutes. Each participant took part in one experimental session. After the experiment, participants filled in a questionnaire to assess strategies and other factors that may have influenced performance.

4.3.4 EEG recording

The EEG was recorded continuously using BrainAmp amplifiers (Brain Products, Munich, Germany) from 64 Ag/AgCl electrodes (actiCAP) positioned according to the international modified 10-20 system. Vertical (vEOG) and horizontal electrooculograms (hEOG) were recorded as voltage difference between electrodes positioned above and below the left eye, and to the left and right canthi of the eyes, respectively. All channels were initially referenced to FCz and re-referenced offline to the average of all electrodes. Electrode impedances were kept below 5kΩ. The sampling rate was 1000 Hz with a high cut-

off filter of 250 Hz (half-amplitude cut-off, 30dB/oct) and a low cut-off filter of 0.016 Hz (half-amplitude cut-off, 6dB/oct).

4.3.5 Data analysis

Analysis was performed with custom MATLAB code (Mathworks, Natick, MA), utilizing functions from the open-source Fieldtrip toolbox (http://www.ru.nl/fcdonders/fieldtrip/).

Behavioral data

Incorrect choices, defined as the reporting of a color/line-orientation pair not present in the display (see **Procedure**), were discarded from all further analyses (5.3 ± 0.12 %). To identify at which trial during a block a participant showed statistically reliable learning of the current value rule, we analyzed the trial-by-trial choice dynamics using the state-space framework introduced by Smith and Brown (2003), and implemented by Smith et al. (2004). This framework entails a state equation that describes the internal learning process as a hidden Markov or latent process and is updated with each trial. The learning state process estimates the probability of a high value choice in each trial and thus provides the learning curve of participants (Figure 4.2B, 4.3B). The algorithm estimates learning from the perspective of an ideal observer that takes into account all trial outcomes of participants' choices in a block of trials to estimate the probability that the outcome in a single trial is a high value response or a low value response. This probability is then used to calculate the confidence range of observing a high value response. We defined the learning trial (T_{Learn}) as the earliest trial in a block at which the lower confidence bound of the probability for a high value response exceeded the $p = 0.5$ chance level. This corresponds to a 0.95 confidence level for an ideal observer to identify learning. The very first block of an experimental session and blocks in which no learning was identified were removed from further analyses. For most analyses, trials were split based on their occurrence prior to T_{Learn} and after, into *during* learning trials and *after* learning trials, respectively.

Reaction times *during* and *after* learning were compared by computing the difference between *during* and *after* learning trials per participant and comparing these differences across participants against a distribution with a mean of zero (t-test, $\alpha = 0.05$). We tested whether reaction time and the probability to switch between colors were dependent on the previous trials' feedback. For the first analysis, we simply compared reaction time in trial n for trials in which trial n-1 ended with a high value feedback and trials in which trial n-1 ended with a low value feedback. This comparison was made within participants as well as between participants (t-test, $\alpha = 0.05$). For the latter analysis, we extracted trials in which a switch of choice was made from stimulus 1 (color 1) to stimulus 2 (color 2) or vice versa. We then computed a ratio of switch trials that followed a low value feedback versus switch trials that followed a high value feedback. This ratio was then compared across participants against a distribution with a mean of one (t-test, $\alpha = 0.05$).

EEG data

For the N2pc analysis the EEG data was segmented into epochs of 700ms, starting 200ms prior to stimulus display onset and ending 500ms after stimulus display onset. The time period from -200 to 0 ms was used as baseline. Trials with blinks (vEOG > 80 μV) or horizontal eye movements (hEOG > 40 μV) were excluded from all analyses (across participants: 78±17 trials). The total trial number available for analysis following artifact removal and removal of non-learned blocks (see Methods above) was 1047±25 trials across participants. The N2pc was measured at parieto-occipital electrode sites (PO3/4, PO7/8) as lateralized response to the laterally presented colored stimulus. The choice of electrode locations was based on the previously shown topography of N2pc subcomponents (Hickey et al., 2009) and equivalent to earlier studies (Feldmann-Wüstefeld and Schubö, 2013; Feldmann-Wüstefeld et al., 2015). Difference waves were calculated by subtracting activity ipsilateral from activity contralateral to the lateral stimulus, and averaged separately for chosen and non-chosen stimuli to isolate choice-related N2pc differences *during* and *after* learning. In line with previous studies, mean amplitudes for the N2pc were computed for

the time interval from 200 to 300 ms after stimulus display onset (Luck and Hillyard, 1994b; Hickey et al., 2010; Woodman and Luck, 1999; Kiss et al., 2008). Initial comparisons were made using a two-way repeated-measures ANOVA with the factors selection (chosen vs. non-chosen) and learning (during vs. after), and followed up by one-way repeated-measures ANOVAs ($\alpha = 0.05$) with the factor learning conducted separately for chosen and non-chosen stimuli.

To investigate whether feedback in trial n-1 had an impact on the mean amplitude of the N2pc component in trial n, we isolated trials in which a choice to a lateralized color target followed a choice to the same color target presented lateralized, to verify that any effect was solely due to feedback. Specifically, a trial combination (trial n and n-1) was only selected for analysis if, e.g., a response was made to color 1 in trial n-1 and in trial n, and stimulus color 1 was presented at a lateralized position (left or right) in both trials. Following this restriction, total trial numbers available for this analysis across participants were 212±7. These trials were then sorted based on the feedback (high or low) received in trial n-1 and the mean amplitude of the N2pc component in trial n was compared in these two groups of trials. Comparisons were made using two-way repeated-measures ANOVA with the factors feedback value in trial n-1 and learning.

For FRN component analyses we extracted the data into 800ms epochs, lasting from -200 to 600ms around the feedback event. Similar to previous studies (e.g. Hajcak et al., 2006; Cohen et al., 2007), the FRN component was isolated at the Fz electrode, and as a control additionally at the FCz electrode (see Results section, data not shown). Difference waves were calculated by subtracting activity for high value feedback from activity for low value feedback in the 250 – 325 ms following feedback onset, which generally fell within the time range investigated in previous studies (see for review: Walsh and Anderson, 2012). The comparison of FRN mean amplitude *during* versus *after* learning was computed using two-way repeated measures ANOVA, with the factors feedback value and learning.

To assess a more general effect of learning on feedback processing independent of valence (FRN), we performed a 3-way ANOVA with the factors learning, feedback value, and time window (twelve non-

overlapping 50-ms windows ranging from 0-600ms post feedback). Follow-up tests of simple main effects were done using one-way ANOVA's in each time window with p-values corrected for multiple comparisons using the Bonferroni-Holm method.

For visualization purposes only, the N2pc and FRN displayed in Figures 4.4-4.6 were smoothed with a moving average filter of 25ms (40Hz).

Correlations

We compared mean differences in N2pc amplitudes following low versus high value feedback with mean behavioral measures on an individual participant level using Pearson correlation ($\alpha = 0.05$). The three behavioral measures tested included i) the proportion of blocks successfully learned, ii) the mean trial at which learning was deemed successful across blocks (i.e. T_{Learn}), and iii) mean reaction time. These three behavioral measures were not correlated across participants (Pearson correlation, all $p > 0.05$). We compared correlation coefficients obtained e.g. *during* learning versus *after* learning by using a z-test to assess differences between two dependent correlations (Steiger, 1980). When the observed z-value was greater than |1.96|, we considered the correlation coefficients significantly different.

4.4 Results

4.4.1 Reversal learning

Behavioral results are plotted in Figure 4.3. Participants performed the task very well and generally showed quick increases in the proportion of high value choices following a value reversal (Figure 4.2B, Figure 4.3Ai, Bi), in line with the behavioral assumptions (Figure 4.1A). This is further shown in the distribution of block lengths observed across all participants, whereby the majority of blocks had a length of approximately 25 trials, indicating a performance of 80% high value choices around the time trial 25

was reached (Figure 4.2Aii, note that blocks shorter than 25 trials were possible following the rejection of incorrect responses, see Methods). Participants performed a mean of 41.1 blocks per experimental session. 86% out of all blocks were successfully learned (across participants: 86 ± 1.6%). Learning of the current value rule across blocks and participants occurred within a mean of 12.5 ± 0.5 trials as identified using the Expectation Maximization algorithm by Smith et al. (2003) (Figure 4.3B, median: 11).

Reaction time significantly decreased with learning of the current value rule. Participants showed on average 13.2 ms shorter reaction times in trials after acquiring the current rule (*after* learning, 609±10ms) as opposed to trials beforehand (*during* learning, 622±12ms) ($t_{(20)} = 3.31$, $p = .004$, Figure 4.3C).

4.4.2 Attention deployment changes with learning

Mean amplitudes for the N2pc were computed for the time interval from 200 to 300 ms after stimulus display onset (Figure 4.4). An initial 2-way repeated-measures ANOVA tested the effects of the factors selection and learning on N2pc amplitudes. A main effect of selection ($F_{1,20} = 44.04$, $p < .001$) showed that a pronounced N2pc was elicited when the chosen stimulus was presented at a lateralized position ($\Delta_{(contra\text{-}ipsi)} = -1.901 \pm 0.29$ μV), which was substantially reduced when the non-chosen stimulus was presented laterally and the chosen stimulus on the vertical midline ($\Delta_{(contra\text{-}ipsi)} = -0.271 \pm 0.16$ μV). A main effect of learning additionally suggested that N2pc amplitudes differed *during* learning and *after* learning ($F_{1,20} = 10.79$, $p = .004$). The post-hoc comparison showed that N2pc amplitudes were significantly larger (more negative) *during* learning ($\Delta_{(contra\text{-}ipsi)} = -1.155 \pm 0.20$ μV) than *after* learning ($\Delta_{(contra\text{-}ipsi)} = -1.016 \pm 0.20$ μV). Therefore, as initially predicted (Figure 4.1B) we found main effects of both learning and selection on the N2pc. However, contrary to our expectation (Figure 4.1B), we did not find a significant interaction between the two ($F_{1,20} = 1.04$, $p = .319$).

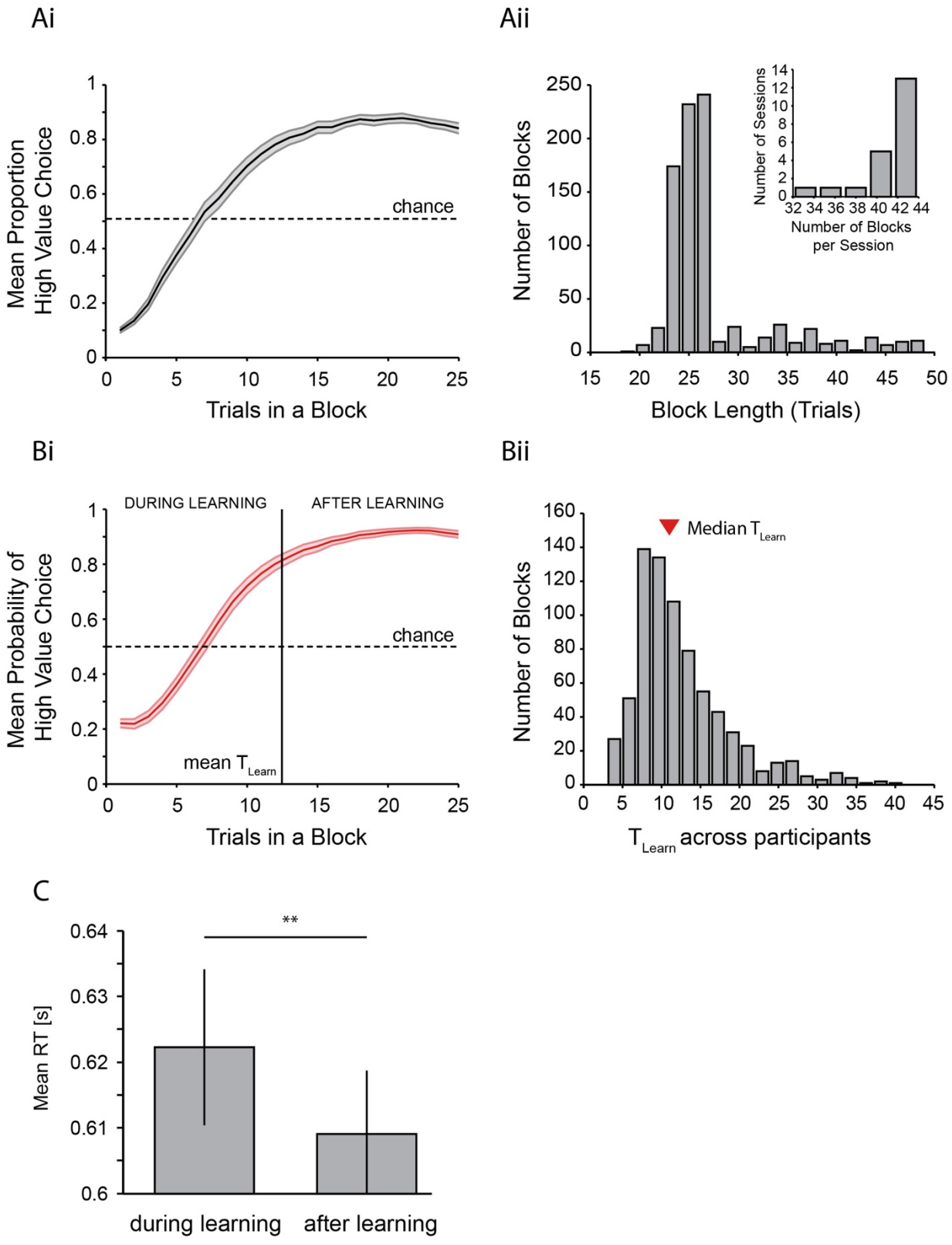

Figure 4.3 Behavioral performance of the reversal learning task.

(A) shows performance across all participants with the mean proportion of high value choices across trials (Ai) and a histogram of the length of blocks (Aii) as well as numbers of blocks per session across all participants (Aii inset). Length of blocks can occasionally be shorter than the minimum of 25 trials because incorrect or late responses were not counted towards block lengths. **(B)** shows performance as measured using the EM algorithm. The mean probability of a high value choice with the average trial at which learning has occurred across participants is shown in Bi, and the distribution of learning trials across all blocks of all participants is shown in Bii (mean learning trial: 12.51; median learning trial: 11). **(C)** displays the effect of learning on reaction time (RT) across participants. Two asterisks indicate $p < 0.01$ (t-test on RT difference (*during-after* learning) across participants).

While the absolute magnitude of the N2pc for the non-chosen stimulus was higher *during* learning than *after* learning (Figure 4.4Cii), and this was significant as a main effect of a one-way ANOVA ($F_{1,20} = 5.36$, $p = .031$) as predicted (Figure 4.1B, right), the magnitude for the chosen stimulus was virtually identical *during* and *after* learning ($F_{1,20} = 1.41$, $p = 0.249$, Figure 4.4Ci), which does not match our prediction (Figure 4.1B, left). Thus, our results suggest that the primary effect of learning on the N2pc amplitude in our task is to suppress attention to non-chosen distractors, rather than to enhance attention to chosen targets. The lack of apparent target enhancement might explain why the two-way interaction was not significant, despite the apparent effect of learning on the non-chosen stimulus. Given this lack of a significant interaction, the differential results should be treated as suggestive, rather than definitive.

In summary, N2pc results showed that attention was mainly deployed to the chosen stimulus compared with the non-chosen stimulus, and that attention deployment was generally more pronounced *during* learning compared with *after* learning. In contrast to our hypotheses (Figure 4.1B), the direction of the effects of learning were not opposing for chosen and non-chosen stimuli. Thus, even though we did not observe an interaction between learning and selection, we nevertheless found evidence suggestive of successful learning mainly leading to a decrease in processing of the non-chosen compared with the chosen stimulus (Figure 4.4C). We therefore found partial evidence in line with our hypothesis for the effect of learning on non-chosen stimuli (Figure 4.1B right).

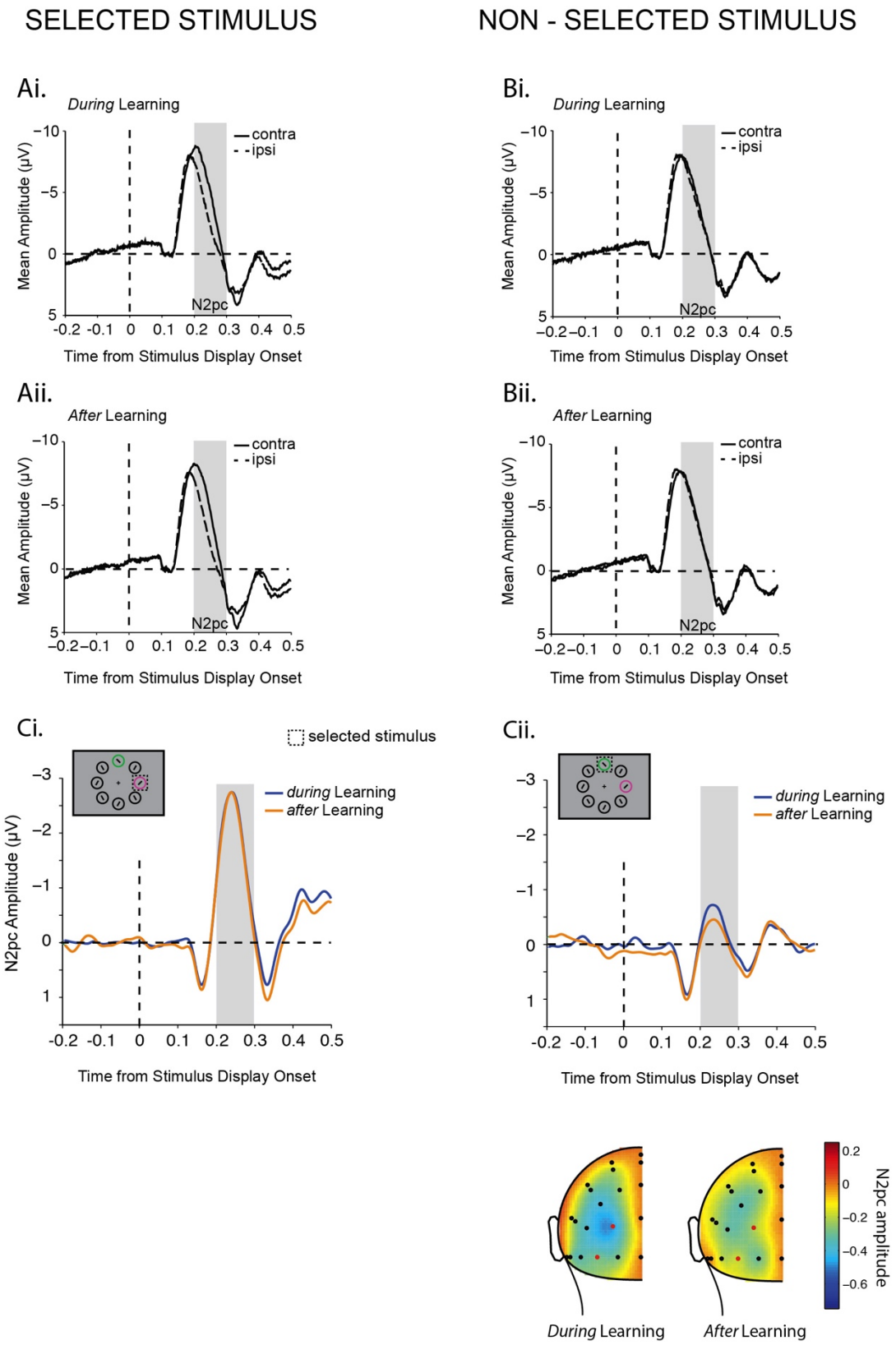

Figure 4.4 Lateralized N2pc components.

(A-B) Contra- and ipsi-lateral mean amplitudes from pooled left (PO3, PO7) and right (PO4, PO8) parieto-occipital electrodes are shown aligned to target display onset for lateralized chosen (A) and non-chosen **(B)** stimuli *during* (Ai, Bi) and *after* (Aii, Bii) learning. The grey bars highlight the N2pc time window of analysis (200-300ms). N2pc difference waves contrasted *during* **(Ci)** and *after* **(Cii)** learning. Example trials are illustrated in the top left corners. The topography of the N2pc (200-300ms) for non-chosen stimuli is shown below Cii.

4.4.3 Low value feedback is followed by increased attentional target selection

To investigate the impact of low or high value feedback on behavioral or electrophysiological measures, we assessed whether reaction time, probability to switch color choices, and N2pc mean amplitudes were affected by the previous trial's feedback. Whether the feedback in trial n-1 was of high or low value had no impact on reaction time in the following trial n (across all trials: $t_{(12,470)}$= -0.36, p = .719). It did, however, have an impact on the likelihood to switch choices from stimulus color 1 to stimulus color 2 or vice versa. Across participants, a choice switch was more likely to occur following a low value feedback compared with a high value feedback ($t_{(20)}$ = 5.10, p = 5.4x10^{-5}).

Feedback in trial n-1 also had an impact on N2pc mean amplitude in trial n. Overall, if a choice was made to the same target in trials n-1 and n (Figure 4.5A), the N2pc amplitude in trial n was larger following a low value feedback in trial n-1 ($\Delta_{(contra-ipsi)}$ = -2.041 ± 0.30 μV), compared with following a high value feedback in trial n-1 ($\Delta_{(contra-ipsi)}$ = -1.671 ± 0.33 μV) ($F_{1,20}$ = 4.52, p = .046) (Figure 4.5B). This was also the case when we did not explicitly control for the choice in trial n (i.e. a choice switch could occur from trial n-1 to trial n, $F_{1,20}$ = 5.51, p = .029, data not shown). We were interested in whether feedback had a differential effect on N2pc amplitudes depending on the current state of learning, and therefore separated trials into *during* learning and *after* learning trials. We did not find a significant interaction between the factors feedback and learning in a two-way ANOVA ($F_{1,20}$ = 0.82, p = .375). Nevertheless, the difference in N2pc amplitudes following high versus low value feedback tended to be greater *during* learning than *after* learning (Figure 4.5C). As previously, this should be treated as suggestive rather than definitive.

Across individual participants, this difference in N2pc amplitude following high versus low value feedback was significantly correlated with learning performance *during* learning (Figure 4.5D left), but

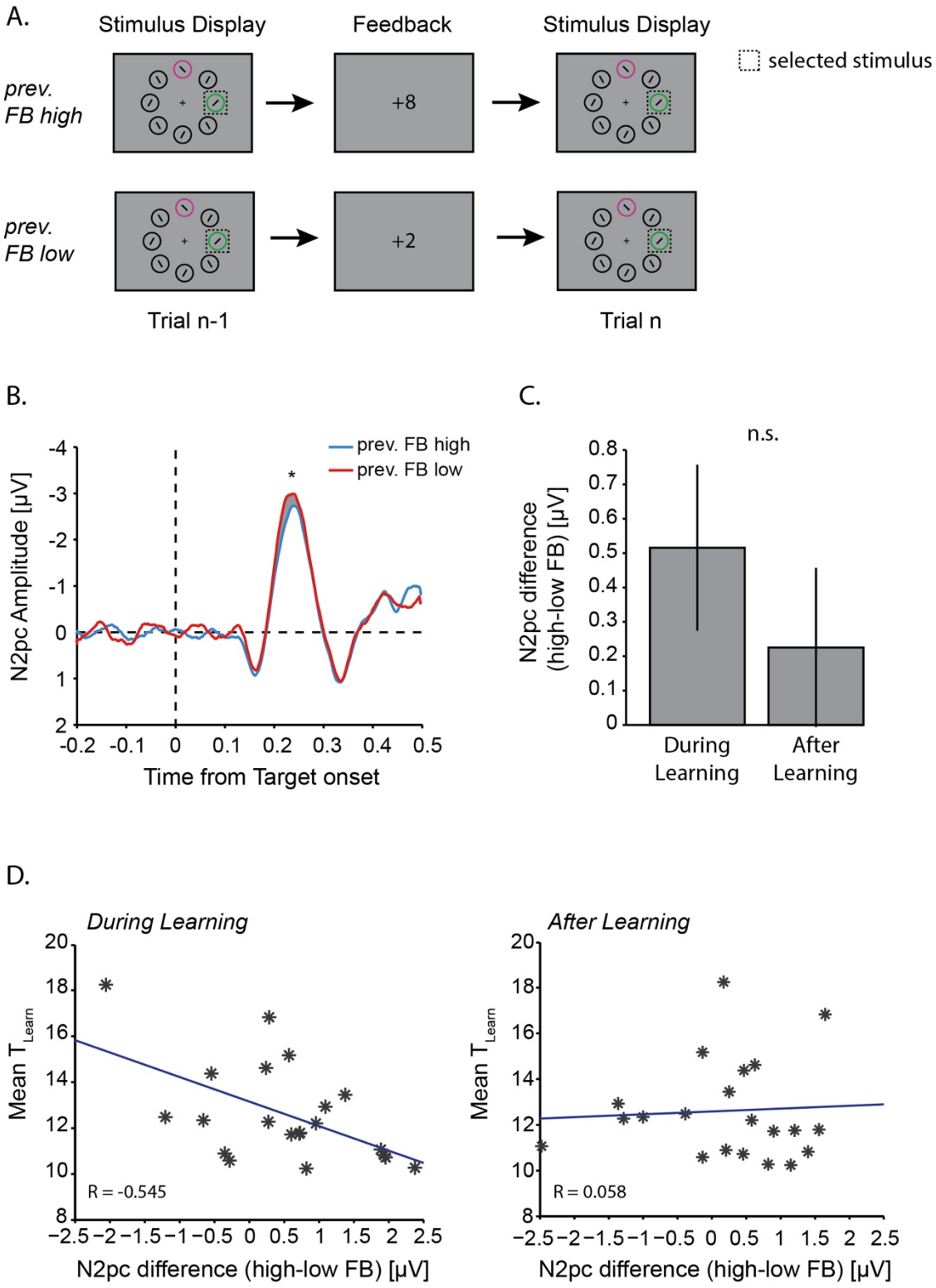

Figure 4.5 Effect of previous trial feedback on N2pc amplitude.

(A) Illustration of example trial sequences for 'previous feedback high' and 'previous feedback low' trials. Trials for analysis were chosen based on the previous trial feedback (n-1), the N2pc analysis was done on trial n. **(B)** N2pc difference wave in trial n following high value (blue line) or low value (red line) feedback in the previous trial. Grey shaded area highlight the N2pc analysis window (200-300ms after stimulus display onset). Asterisk indicates $p \leq 0.05$ (1-way ANOVA). **(C)** Mean difference in N2pc amplitude following high versus low value feedback *during* learning and *after* learning. **(D)** Correlation between individual participants' T_{Learn} and N2pc amplitude differences between feedback (previous high-value feedback – previous low-value feedback) *during* learning (left) and *after* learning (right). Blue lines represent least square fit. Note that large positive differences indicate a large N2pc difference for high versus low-value feedback in trial n-1.

not *after* learning (Figure 4.5D right). Specifically, the greater the individual difference in N2pc amplitude following high versus low value feedback *during* learning, the faster the individual learned, i.e. T_{Learn} was smaller (Pearson correlation, $R = -0.545$, $p = .011$) (Figure 4.5D left). However, the difference in N2pc amplitudes following high versus low value feedback *after* learning was not related to learning performance (Pearson correlation, $R = 0.058$, $p = .803$) (Figure 4.5D right). This difference in correlation coefficients between *during* learning and *after* learning was significant (z-test to compare R-values, $z = 2.12$, $p = .034$). The difference in N2pc amplitude following high versus low value feedback was not correlated with average reaction time or the proportion of blocks learned (all $p > .05$).

4.4.4 Feedback processing is increased *during* learning

Considering the finding that feedback has a differential effect on N2pc amplitudes, and that this effect specifically *during* learning correlates with successful learning behavior, we asked whether feedback processing was affected by learning. We therefore computed the mean amplitude of the feedback-related negativity (FRN) as a proxy for negative feedback processing (Miltner et al., 1997). The FRN, computed as the difference between low and high value feedback presentation, was measured at the Fz electrode, since the amplitude difference between low and high value feedback was largest at this electrode. The qualitative and quantitative pattern of results did not change when the FRN was measured at the FCz electrode instead (data not shown). The analysis was done within the 250 – 325ms after feedback presentation, since the difference between low and high value was largest in this window (*see* below) and

it generally fell within the range used in the literature (see for review: Walsh and Anderson, 2012). We found that processing of feedback (low and high value) was generally increased *during* learning compared with *after* learning (Figure 4.6A), and the difference between low and high value feedback (FRN) was more pronounced *during* learning compared with *after* learning (*during*: $\Delta_{(lowFB\text{-}highFB)}$ = -0.791 ± 0.17 μV; *after*: $\Delta_{(lowFB\text{-}highFB)}$ = -0.426 ± 0.18 μV). The resulting FRN was therefore substantially larger *during* learning as compared with *after* learning (Figure 4.6B). This was confirmed with a two-way ANOVA that showed a significant main effect of feedback value ($F_{1,20}$ = 14.70, p = .001), a significant main effect of learning ($F_{1,20}$ = 37.18, p < .001), and a significant interaction between the two parameters ($F_{1,20}$ = 6.04, p = .023). Topographical maps of the amplitude difference between low and high value feedback *during* and *after* learning are shown in Figure 4.6C.

In addition to the change in FRN amplitude with learning, visual inspection of the plots (Figure 4.6A), revealed a much longer effect of learning that was distinct from the effect of feedback value and the interaction of feedback value and learning. To tease apart these effects and to determine the time range of the effect of learning, we ran a three-way ANOVA with the factors learning (*during* learning, *after* learning), feedback value (high, low), and time window, where we defined twelve 50ms non-overlapping time windows running from 0 to 600ms following feedback onset. We found interactions between the factors learning and time window ($F_{11,220}$ = 7.13, p < .001), and feedback value and time window ($F_{11,220}$ = 3.88, p <. 001). Follow-up simple main effects across time windows showed that feedback processing per se differed with learning in all time windows from 150ms to 400ms following feedback onset (F-values between 18.07 and 47.29, all p < .005, p-values in all other time windows > .05, Bonferroni-Holm multiple comparison corrected). A simple main effect of feedback was only found in the 250-300ms time window ($F_{1,20}$ = 16.39, p = .008, all other p >.05, Bonferroni-Holm multiple comparison corrected), confirming the initial FRN analysis above. The previous suggests that feedback processing independent of valence was increased *during* learning in the time window from 150-400ms following feedback onset (Figure 4.6D).

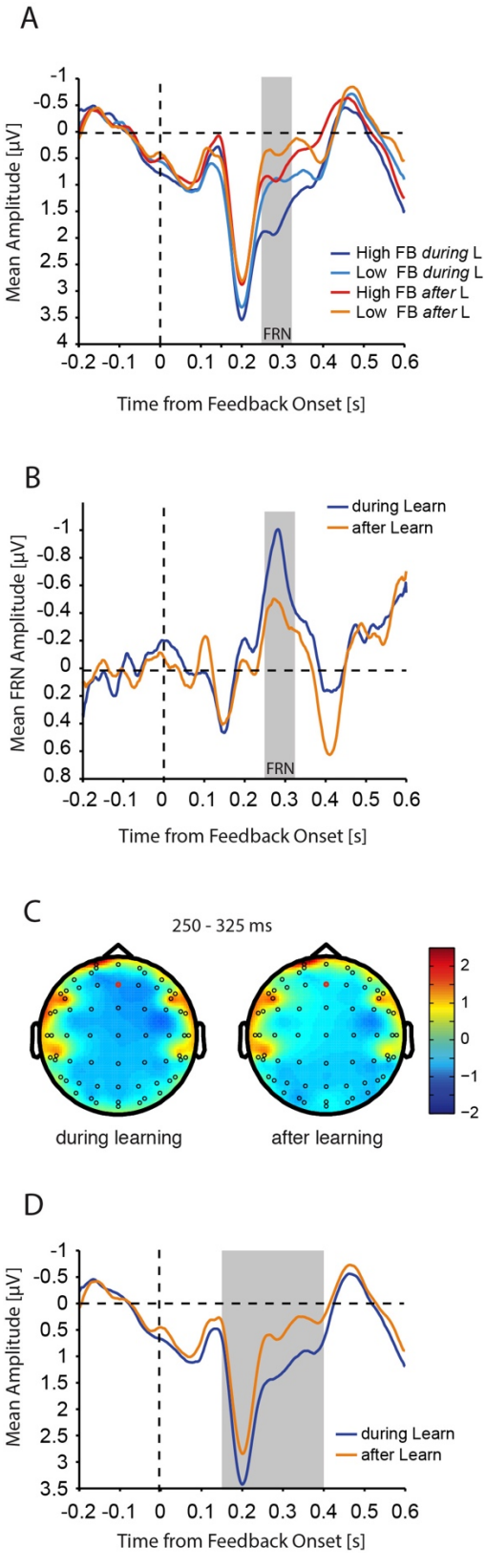

Figure 4.6 FRN amplitudes.

(A) Aligned to feedback presentation onset, mean activity at Fz electrode is shown for high and low value feedback *during* and *after* learning trials. **(B)** The mean difference (low value feedback – high value feedback) for the FRN wave is shown *during* and *after* learning. The grey bar indicates the analysis time window for the FRN (250 – 325 ms after feedback onset). For visualization purposes only, the FRN wave was smoothed with a moving average filter of 5ms. **(C)** Topography of the FRN (difference elicited by low versus high value feedback) *during* and *after* learning. Red circles identify Fz electrodes. **(D)** Mean activity from (A) is shown collapsed across feedback valence to illustrate the time span of the simple main effect of learning on feedback processing (grey bar).

4.5 Discussion

In this study, we implemented a value-based reversal learning task to explore in more detail how attentional target selection and feedback processing is realized in a highly volatile task environment. We measured the N2pc, an EEG component thought to reflect attentional target selection, and the FRN, an EEG component thought to reflect negative feedback processing or prediction error encoding, while participants performed a value-based reversal learning task with probabilistic feedback. Participants were required to frequently adjust their color-based stimulus choice using recent reward feedback. We found that i.) Participants used feedback efficiently to reverse back and forth between the two stimulus choices in accordance with the reversal of their respective values (Figures 4.1A, 4.2, 4.3), ii.) Successful learning of the current value-contingency led to a decrease in N2pc amplitudes, which was particularly evident for non-chosen (distractor) stimuli compare with chosen (target) stimuli (Figure 4.4), iii.) Negative feedback in the previous trial was associated with an increase in N2pc mean amplitude, which was selectively correlated with an enhanced learning rate *during* learning (Figure 4.5), and iv.) FRN amplitudes were increased *during* learning of the current value contingencies, which co-occurred with a more general increase of feedback processing *during* learning (Figure 4.6).

We live in an environment that is usually much more uncertain and volatile than a classical experimental setting, in which objects or actions need to be continuously evaluated for their relevance to our current

goal. We attempted to imitate some of this volatility by employing a value-based reversal learning task, albeit one that is clearly more restrictive than the world outside the laboratory.

To the best of our knowledge this is the first study that investigated learned value-dependent changes of the N2pc amplitude elicited by chosen (i.e., selected) and non-chosen (non-selected) stimuli in a task in which participants were free to select a stimulus for response. Most studies that have investigated the effects of value or reward on attentional stimulus selection in behavior, have used tasks with a predefined, fixed target and distractor-assignment, and in which the trial-by-trial association between the specific stimulus features and reward were in fact irrelevant to solving the task (Della Libera and Chelazzi, 2009; Hickey et al., 2010, 2015, Anderson et al., 2011a, 2013; Itthipuripat et al., 2015; Sawaki et al., 2015; Feldmann-Wüstefeld et al., 2016), often to dissociate reward-related processes from goal-related processes of attentional selection. Or they have used tasks in which value associations of targets or cues were kept constant (Kiss et al., 2009; Raymond and Brien, 2009; Krebs et al., 2010; Le Pelley et al., 2013; San Martin et al., 2016), or if they changed, were specifically trained (Navalpakkam et al., 2010). None of these studies allowed insight into how attentional selection changes when the values of target stimuli change unannounced and require repeated adjustment of behavior.

We used probabilistic feedback so that subjects needed to keep track of feedback over multiple trials to determine the current high-value stimulus. Following an unannounced value reversal, participants tended to continue to choose the low value (previously high value) stimulus for response for multiple trials, before switching their choice behavior to the current high value stimulus. Reaction times were longer *during* learning than *after* learning (Figure 4.3C), suggesting that participants optimized their attention allocation to stimulus features with learning of the current value contingencies.

Since both stimuli were repeatedly associated with a high and a low value, both were frequently selected for a response. Thus, the dissociation between target and distractor is not as clear on a trial by trial basis as in previous literature (see above). For this reason, it was initially unclear to what extent processing of the chosen (that is, the selected target) stimulus would be enhanced throughout learning and how

efficiently the brain could evade attentional capture by the non-chosen (non-selected distractor) stimulus, which always posed a distraction to solving the task quickly and accurately. We initially predicted that as attentional prioritization shifts towards the chosen stimulus with learning, this would concomitantly be reflected by an N2pc increase for the chosen stimulus (Figure 4.1B left) and an N2pc decrease for the non-chosen stimulus (Figure 4.1B right). Instead we found that N2pc amplitudes decreased with learning in general, which was true on average for both chosen and non-chosen stimuli. However, this decrease in amplitude with learning seemed more apparent for the non-chosen stimulus (Figure 4.4C). This suggests that the primary effect of learning in this task was a decrease in attention to the non-chosen lateralized stimulus, which potentially indicates suppression (Figure 4.4Cii). In contrast, the amplitude of the N2pc that reflected processing of the chosen lateralized stimulus did not seem to substantially change with learning (Figure 4.4Ci). Thus, we find some evidence for our initial hypothesis of how learning affects processing of non-chosen stimuli (Figure 4.1B right), but not for our hypothesis of how learning affects processing of chosen stimuli (Figure 4.1B left). These results suggest that following a value reversal, when participants needed to actively re-evaluate their current choices, distraction by the non-chosen stimulus was not as effectively evaded as *after* learning, when participants often showed plateau-performance with a high probability of choosing the currently high-valued stimulus (Figures 4.2, 4.3).

These results suggest that efficient attention allocation in this highly volatile task design was more likely observed for processing of the non-chosen stimulus in form of a reduced or suppressed N2pc, and not as an N2pc enhancement of the chosen stimulus. We should note however that although we observed suppression for lateralized non-chosen stimuli, independent of learning, the amplitude elicited in this time interval was still negative, and not positive as has been observed previously (e.g. Hickey et al., 2009; Feldmann-Wüstefeld et al., 2015, 2016; Sawaki et al., 2015). It is therefore difficult to decide whether value learning has led to a reduced capture by the non-chosen (i.e., non-selected) stimulus or an actual suppression, as both interpretations would account for a reduction in N2pc amplitude. Similarly, it is possible that in a less volatile task design in which learning takes place over a much longer time window

(e.g. days), an effect of learning on attentional processing would predominantly be observed for the chosen (target) stimulus, as has been shown previously (Clark et al., 2015; Itthipuripat et al., 2017). We should therefore be careful of over-interpreting the absence of a strong effect of learning on N2pc amplitudes of the chosen stimulus in this task, as such an effect could have been revealed with a larger number of *after*-learning trials.

That attention and learning are closely intertwined concepts has been the subject of attentional learning theories for some time (Mackintosh, 1975; Pearce and Hall, 1980; Le Pelley, 2004). The Mackintosh and the Pearce and Hall models of attentional learning predict contradicting relationships between attention and learning. According to Mackintosh (1975), attention is biased towards stimuli that have a higher predictive value, as they are more likely to yield a rewarding outcome (e.g. Mackintosh & Little, 1969; Le Pelley et al., 2013). The Pearce and Hall model on the contrary predicts that unexpected and surprising outcomes that lead to a prediction error are associated with an increase in attention (e.g. Wilson et al., 1992; Anderson et al., 2013). Both theories have received extensive empirical support (Pearce and Mackintosh, 2010) and there have been efforts to reconcile their findings (Holland and Gallagher, 1999; Dayan et al., 2000; Le Pelley, 2004; Hogarth et al., 2010). One possible solution suggests that there is a distinction between two aspects of attention in associative learning: attention that is concerned with action, and attention that is concerned with learning. It suggests that one should attend to the most reward-predicting stimuli or features when making a choice, but should attend to the most uncertain stimuli or features when learning from prediction errors (Holland and Gallagher, 1999; Dayan et al., 2000; Hogarth et al., 2010; Gottlieb, 2012).

Even though our task was not designed to address this debate explicitly, we found evidence that in a highly volatile environment that encourages continuous learning, attention is increased during periods of uncertainty in line with Pearce and Hall (1980). In particular, we tested whether feedback itself could influence the allocation of attention to target stimuli and found that low-value feedback was followed by a

larger N2pc amplitude on the next trial than was high-value feedback (Figure 4.5B), when the only difference between two consecutive trials was the feedback received after the first trial (Figure 4.5A). Importantly, only *during* learning did this difference in N2pc amplitude following low versus high value feedback correlate with individual participants' learning rates - participants in which the difference in feedback-dependent N2pc amplitude was particularly large learned faster (Figure 4.5D left). This suggests that *during* learning, when participants needed to actively reevaluate their stimulus choices and relearn the current value contingency, allocating more attention to the choice stimulus after experiencing a negative outcome compared with a positive outcome, led to faster and more successful adjustment of behavior according to the current value contingency. In addition, this suggests that in a highly volatile task environment that requires continuous learning and value updating, with sources of uncertainty found in the inherent reward probability distributions (70%-30%) and in the sudden value reversals (Yu and Dayan, 2005; Payzan-Lenestour and Bossaerts, 2011), participants allocate more attention to an uncertain compared with a more certain choice stimulus, in line with Pearce and Hall (1980). It is possible that in a less volatile environment in which participants have much longer periods of consistent choices, i.e. periods in which participants know the current value contingency with high certainty and learning presumably no longer takes place, we could have observed an increase in attention during target selection, which would be in line with Mackintosh (1975). Although we are using the N2pc amplitude as a proxy for selective attention, which may limit the implications to be drawn, our results are consistent with the idea that when tasks demand states of active learning, attention is increased following uncertain choice outcomes or events, and this correlates with enhanced learning performance.

Another prominent EEG component that has been shown to change during reversal learning is the feedback-related negativity (FRN) (e.g. Chase et al., 2010; von Borries et al., 2013; Donaldson et al., 2016). The FRN has been thought to encode negative feedback, prediction error signals, outcome valence, and behavioral adjustment (Holroyd and Coles, 2002; Cohen and Ranganath, 2007; Bellebaum and

Daum, 2008; Chase et al., 2010; Walsh and Anderson, 2011; von Borries et al., 2013; Donaldson et al., 2016). Using a probabilistic reversal learning paradigm similar to ours, Chase and colleagues (2010) have shown that the FRN amplitude scales with a negative prediction error signal obtained with a reinforcement learning model, whereby the FRN amplitude was largest following a reversal and diminished as a behavioral adjustment approached. Recent evidence using a reversal learning task, in which positive as well as negative outcomes could signal a need for behavioral adjustment and could be equally unexpected, suggests that the FRN may be more related to outcome valence (positive vs. negative) than to expectancy or behavioral adjustment (von Borries et al., 2013; Donaldson et al., 2016). We computed the FRN as the difference wave between the presentation of low-value and high-value feedback, and then compared the difference *during* learning with *after* learning (Figure 4.6B). We found that the FRN was substantially larger *during* learning than *after* learning. Although we should be careful of over-interpreting, since our task confounds the accumulation of negative outcomes with the need for behavioral adjustment, these results suggest that low-value as well as high-value feedback are processed differently during periods of uncertainty (*during* learning) when stimulus-value associations need to be updated (Figure 4.6A), which cannot be solely explained by differences in outcome valence.

In addition to the effect of learning on the FRN amplitudes, we also observed a more general and longer lasting effect of learning on feedback processing that was independent of feedback valence (Figure 4.6D). Feedback processing differed between 150 and 400ms following feedback onset d*uring* learning of current value contingencies compared with *after* learning, which could indicate an increase in feedback processing per se. The FRN is thought to originate in anterior cingulate cortex (ACC) (Hickey et al., 2010), and this prolonged window of differential feedback processing matches the time-resolved activity level of ACC during reward presentation observed previously (Hickey et al., 2010). In our task, enhanced ACC activity, during periods in which participants experience increased levels of uncertainty that potentially require a behavioral adjustment, could potentially reflect the necessity of increased levels of cognitive control (Shenhav et al., 2013, 2016), or increased activity related to the decision of moving

from a state of exploitation to a state of exploration (Kolling et al., 2016). Together with increased attention *during* learning and following negative feedback, these signals may be part of the underlying neural network that drives behavioral adjustment during periods of increased uncertainty that concludes in the switch of stimulus choice.

Overall, we found evidence that during periods of active behavioral adjustment in a changing and volatile task environment, feedback processing of recent choices and attentional processing is amplified and co-occurs with increases in attentional allocation following low-value feedback compared with high-value feedback that possibly promotes increased learning speed. Following successful learning of current value contingencies, during periods of stable behavior, attentional allocation then becomes potentially more efficient by suppressing non-relevant distractor processing. These results provide insight into how changes in attentional prioritization and feedback processing may support flexible and repeated behavioral adjustments in humans.

4.6 References

Anderson BA (2013) A value-driven mechanism of attentional selection. J Vis 13:1–16.

Anderson BA, Laurent PA, Yantis S (2011a) Learned Value Magnifies Salience-Based Attentional Capture. PLoS One 6:e27926.

Anderson BA, Laurent PA, Yantis S (2011b) Value-driven attentional capture. Proc Natl Acad Sci USA 108:10367–10371.

Anderson BA, Laurent PA, Yantis S (2013) Reward predictions bias attentional selection. Front Hum Neurosci 7:262.

Anderson BA, Laurent PA, Yantis S (2014) Value-driven attentional priority signals in human basal ganglia and visual cortex. Brain Res 1587:88–96.

Awh E, Belopolsky AV, Theeuwes J (2012) Top-down versus bottom-up attentional control: A failed theoretical dichotomy. Trends Cogn Sci 16:437–443.

Bellebaum C, Daum I (2008) Learning-related changes in reward expectancy are reflected in the feedback-related negativity. Eur J Neurosci 27:1823–1835.

Bucker B, Theeuwes J (2017) Pavlovian reward learning underlies value driven attentional capture. Attention, Perception, Psychophys:1–14.

Chase HW, Swainson R, Durham L, Benham L, Cools R (2010) Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. J Cogn Neurosci 23:936–946.

Clark K, Appelbaum LG, van den Berg B, Mitroff SR, Woldorff MG (2015) Improvement in Visual Search with Practice: Mapping Learning-Related Changes in Neurocognitive Stages of Processing. J Neurosci 35:5351–5359.

Cohen MX, Elger CE, Ranganath C (2007) Reward expectation modulates feedback-related negativity and EEG spectra. Neuroimage 35:968–978.

Cohen MX, Ranganath C (2007) Reinforcement learning signals predict future decisions. J Neurosci 27:371–378.

Cohen MX, Wilmes K, van de Vijver I (2011) Cortical electrophysiological network dynamics of feedback learning. Trends Cogn Sci 15:558–566.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 22:4563–4567.

Corbetta M, Shulman GL (2002) Control of Goal-Directed and Stimulus-Driven Attention in the Brain. Nat Rev Neurosci 3:215–229.

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3:1218–1223.

Della Libera C, Chelazzi L (2006) Visual selective attention and the effects of monetary rewards. Psychol Sci 17:222–227.

Della Libera C, Chelazzi L (2009) Learning to attend and to ignore is a matter of gains and losses. Psychol Sci 20:778–784.

Della Libera C, Perlato A, Chelazzi L (2011) Dissociable effects of reward on attentional learning: From passive associations to active monitoring. PLoS One 6:2–7.

Donaldson KR, Ait Oumeziane B, Hélie S, Foti D (2016) The temporal dynamics of reversal learning: P3 amplitude predicts valence-specific behavioral adjustment. Physiol Behav 161:24–32.

Eimer M (2014) The neural basis of attentional control in visual search. Trends Cogn Sci 18:526–535.

Eimer M, Grubert A (2014) Spatial attention can be allocated rapidly and in parallel to new visual objects. Curr Biol 24:193–198.

Feldmann-Wüstefeld T, Brandhofer R, Schubö A (2016) Rewarded visual items capture attention only in heterogeneous contexts. Psychophysiology 53:1063–1073.

Feldmann-Wüstefeld T, Schubö A (2013) Context homogeneity facilitates both distractor inhibition and target enhancement. J Vis 13:1–12.

Feldmann-Wüstefeld T, Uengoer M, Schubö A (2015) You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. Psychophysiology 52:1483–1497.

Gottlieb J (2012) Attention, learning and the value of information. Neuron 76:281–295.

Hajcak G, Moser JS, Holroyd CB, Simons RF (2006) The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. Biol Psychol 71:148–154.

Hickey C, Chelazzi L, Theeuwes J (2010) Reward changes salience in human vision via the anterior cingulate. J Neurosci 30:11096–11103.

Hickey C, Chelazzi L, Theeuwes J (2011) Reward has a residual impact on target selection in visual search, but not on the suppression of distractors. Vis cogn 19:117–128.

Hickey C, Di Lollo V, McDonald JJ (2009) Electrophysiological indices of target and distractor processing in visual search. J Cogn Neurosci 21:760–775.

Hickey C, Kaiser D, Peelen M V (2015) Reward Guides Attention to Object Categories in Real-World Scenes. J Exp Psychol Gen 144:264–273.

Hogarth L, Dickinson A, Duka T (2010) Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal, and uncertainty. In: Attention and Associative Learning. From Brain to Behavior (Mitchell CJ, Le Pelley ME, eds), pp 71–97. Oxford University Press.

Holland PC, Gallagher M (1999) Amygdala circuitry in attentional and representational processes. Trends Cogn Sci 3:65–73.

Holroyd C, Coles M (2002) The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679–709.

Hopf JM, Luck SJ, Boelmans K, Schoenfeld MA, Boehler CN, Rieger J, Heinze HJ (2006) The Neural Site of Attention Matches the Spatial Scale of Perception. J Neurosci 26:3532–3540.

Hopf JM, Luck SJ, Girelli M, Hagner T, Mangun GR, Scheich H, Heinze HJ (2000) Neural sources of focused attention in visual search. Cereb cortex (New York, NY 1991) 10:1233–1241.

Irons JL, Leber AB (2016) Choosing attentional control settings in a dynamically changing environment. Attention, Perception, Psychophys:1–18.

Itthipuripat S, Cha K, Byers A, Serences JT (2017) Two different mechanisms support selective attention at different phases of training. PLoS Biol. 15:e2001724.

Itthipuripat S, Cha K, Rangsipat N, Serences JT (2015) Value-based attentional capture influences context-dependent decision-making. J Neurophysiol 114:560–569.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315–341.

Kiss M, Van Velzen J, Eimer M (2008) The N2pc component and its links to attention shifts and spatially selective visual processing. Psychophysiology 45:240–249.

Kiss M, Driver J, Eimer M (2009) Reward priority of visual target singletons modulates ERP signatures of attentional selection. Psychol Sci 20:245–251.

Kolling N, Wittmann MK, Behrens TEJ, Boorman ED, Mars RB, Rushworth MFS (2016) Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci 19:1280–1285.

Krebs RM, Boehler CN, Woldorff MG (2010) The influence of reward associations on conflict processing in the Stroop task. Cognition 117:341–347.

Le Pelley ME (2004) The role of associative history in models of associative learning: a selective review and a hybrid model. Q J Exp Psychol B 57:193–243.

Le Pelley ME, Vadillo M, Luque D (2013) Learned predictiveness influences rapid attentional capture: evidence from the dot probe task. J Exp Psychol Learn Mem Cogn 39:1888–1900.

Luck SJ, Hillyard SA (1994a) Electrophysiological correlates of feature analysis during visual search. Psychophysiology 31:291–308.

Luck SJ, Hillyard SA (1994b) Spatial filtering during visual search: evidence from human electrophysiology. J Exp Psychol 20:1000–1014.

Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev 82:276–298.

Mackintosh NJ, Little L (1969) Intradimensional and extradimensional shift learning by pigeons. Psychon Sci 14:5–6.

Miltner WHR, Braun CH, Coles MGH (1997) Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a "Generic" Neural System for Error Detection. J Cogn Neurosci 9:788–798.

Munneke J, Hoppenbrouwers SS, Theeuwes J (2015) Reward can modulate attentional capture, independent of top-down set. Atten Percept Psychophys 77:2540–2548.

Navalpakkam V, Koch C, Rangel A, Perona P (2010) Optimal reward harvesting in complex perceptual environments. Proc Natl Acad Sci U S A 107:5232–5237.

Payzan-Lenestour E, Bossaerts P (2011) Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput Biol 7.

Pearce J, Hall G (1980) A model for Pavlovian learning: Variation in the effectiveness of conditioned but not unconditioned stimuli. Psychol Rev 87:532–552.

Pearce J, Mackintosh NJ (2010) Two theories of attention: a review and possible integration. In: Attention and Associative Learning. From Brain to Behavior (Mitchell CJ, Le Pelley ME, eds), pp 11–39. Oxford University Press.

Posner M, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci 13:25–42.

Raymond JE, O'Brien JL (2009) Selective Visual Attention and Motivation. 20:981–989.

Sali AW, Anderson BA, Yantis S (2014) The role of reward prediction in the control of attention. J Exp Psychol Hum Percept Perform 40:1654–1664.

San Martin R, Appelbaum LG, Huettel SA, Woldorff MG (2016) Cortical Brain Activity Reflecting Attentional Biasing Toward Reward-Predicting Cues Covaries with Economic Decision-Making Performance. Cereb Cortex 26:1–11.

Sawaki R, Luck SJ, Raymond JE (2015) How attention changes in response to incentives. J Cogn Neurosci 27:2229–2239.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Shenhav A, Cohen JD, Botvinick MM (2016) Dorsal anterior cingulate cortex and the value of control. Nat Neurosci 19:1286–1291.

Smith AC, Brown EN (2003) Estimating a State-Space Model from Point Process Observations. Neural Comput 15:965–991.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki W a, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. J Neurosci 24:447–461.

Steiger JH (1980) Tests for comparing elements of a correlation matrix. Psychol Bull 87:245–251.

von Borries AKL, Verkes RJ, Bulten BH, Cools R, de Bruijn ERA (2013) Feedback-related negativity codes outcome valence, but not outcome expectancy, during reversal learning. Cogn Affect Behav Neurosci 13:737–746.

Walsh MM, Anderson JR (2011) Modulation of the feedback-related negativity by instruction and experience. Proc Natl Acad Sci USA 208:19048–19035.

Walsh MM, Anderson JR (2012) Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neurosci Biobehav Rev 36:1870–1884.

Wilson PN, Boumphrey P, Pearce JM (1992) Restoration of the orienting response to a light by a change in its predictive accuracy. J Exp Psychol 44:17–36.

Womelsdorf T, Everling S (2015) Long-Range Attention Networks: Circuit Motifs Underlying Endogenously Controlled Stimulus Selection. Trends Neurosci 38:682–700.

Woodman GF, Luck SJ (1999) Electrophysiological measurement of rapid shifts of attention during visual search. Nature 400:867–869

Woodman GF, Luck SJ (2003) Serial deployment of attention during visual search. J Exp Psychol Hum Percept Perform 29:121–138.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.

Chapter 5

General Discussion

5.1 Summary

Attention provides the gateway between the mass of information that surrounds us, and the small subset of that information that we select for further processing, action or learning. In this thesis, I have investigated some of the neural mechanisms that underlie successful attentional deployment, when the attentional target is cued to the subject (**Chapter 2**), and when the attentional target must be learned (**Chapter 3, 4**), in macaque monkeys (**Chapter 2, 3**) and in humans (**Chapter 4**). I have shown in **Chapter 2** that neurons in anterior cingulate cortex (ACC) and dorsal prefrontal cortex (PFC) of the macaque brain correlate their spiking activity during periods of covert shifts of attention to a cued target stimulus. These inter-areal spiketrain correlations were selectively evident for neuron pairs in which both neurons preferred contra- over ipsi-lateral attention shifts, indicating that the spiketrain correlations observed carried specific spatial attention information and therefore likely supported successful shifts of attention in light of distractors. In **Chapter 3**, I have identified feature-specific prediction error signals in ACC and PFC, as well as in the caudate nucleus (CD) and ventral striatum (VS) of the macaque brain, that may aid in learning of the current attentional targets in a volatile and changing reversal learning environment. In particular, neurons in ACC, CD and VS that were informative about the positive or negative valence of the outcome and the immediately preceding color choice, could potentially serve as selective learning signals and aid in assigning credit of an outcome to the source of that outcome in the preceding choice. In **Chapter 4**, I measured two event-related potentials in the human brain – the FRN, potentially an index of prediction errors, and the N2pc, an index of attentional target selection, during the performance of a value-based reversal learning task. I showed that the feedback-related negativity (FRN) is selectively increased during periods of increased uncertainty following a value reversal, and that this co-occurred with increases in N2pc amplitudes in trials following negative feedback, suggesting enhanced

attentional allocation after experiencing uncertain outcomes. The previous chapters allow insights from different angles into the neural mechanisms that support successful attentional deployment during goal-directed behavior.

5.2 In light of attention mechanisms

In the introduction, we learned about three potentially distinct mechanisms of attention – attention for action (or choice), attention for learning, and attention for liking (Gottlieb, 2012). The previous three chapters have allowed us to gain some insight into the former two of these proposed attention mechanisms. When attention is required for accurate action selection, attention should be maximal towards the stimulus or cue with the highest predictability (Mackintosh, 1975). In the cued attention task employed in **Chapter 2**, the stimulus with the highest predictability was the stimulus that matched the color of the fixation cue and the response action was determined based on that stimulus. This was a non-volatile task and the animals had learned that the stimulus that matched the color of the fixation cue was the best predictor of reward outcome and thereby selectively attended to that stimulus to determine the response action. Any uncertainty in this task likely emerged from errors that were due to e.g. lapses in attention, or the inability to withhold a response leading to premature responses. The results from **Chapter 2** indicate that selective spiketrain correlations between neurons in ACC and PFC potentially support successful attention shifts to the stimulus with the highest reward predictability.

In **Chapters 3** and **4**, the environments monkeys and humans were placed in were more volatile and the most reliable predictors for a reward outcome changed frequently. In these environments, it is possible that the 'attention for action' and 'attention for learning' mechanisms dominated at different times, such that after successful learning of the current reward contingencies, in periods of stable and high performance, the 'attention for action' mechanism was in place and attention was successfully shifted to the currently high-valued target, the most reliable predictor for a high reward outcome. However, following a reversal of the high-valued stimulus, an accumulation of low-value outcomes required an

adjustment in behavior – learning – that likely co-occurred with increases in attention to the now uncertain stimuli in the environment. We find evidence for this in **Chapter 4**, where attentional target selection was increased in trials following negative feedback, and in particular during periods of behavioral adjustment immediately following a value reversal. The increased levels of uncertainty arising from value reversals are potentially driven by feature-specific prediction error signals in the fronto-striatal circuit (**Chapter 3**) that signal violations of expectation tied to the specific features selected in the preceding choice and that are largest following a value reversal (Figure 3.4B). I speculate that prediction error signals, like the ones we observe in ACC in **Chapter 3** (Figures 3.5, 3.6, 3.7) underlie an FRN signal observed in **Chapter 4** (Miltner et al., 1997; Hyman et al., 2017), which was greater following a value reversal (Figure 4.6), and in turn may drive the attentional adjustments observed over visual cortex in the following trials to allow for optimal behavioral adjustments when faced with increased uncertainty (Figure 4.5). It is likely that both 'attention for action' and 'attention for learning' are mechanisms that facilitate the detection of uncertainty in the environment, in order to find solutions for reducing this uncertainty (Dayan and Daw, 2008; Gottlieb, 2012).

5.3 In light of uncertainty

Any source of information from the environment can be associated with an uncertainty that can be considered as expected or unexpected uncertainty from the view point of the subject (Yu and Dayan, 2005). Expected uncertainty for instance describes risk, or the inherent stochasticity of the environment that remains even after all contingencies are known (Payzan-Lenestour and Bossaerts, 2011). Unexpected uncertainty in turn results from gross changes in the environment that cause a large violation of expectation (Yu and Dayan, 2005). In **Chapter 4**, expected and unexpected uncertainty can be dissociated easily, with expected uncertainty arising from the inherent reward probability, which was set at 70% when correct (high-value) choices were made, while unexpected uncertainty arose from the sudden reversals in value contingencies (Figure 4.2). In **Chapters 2** and **3**, the dissociation is less straight

forward. Animals in both tasks rarely performed at ceiling levels; indeed, in the cued attention task in Chapter 2, only 60-70% of responses were correct even though contingencies were presumably fully known. The rest of the responses constituted early responses, late responses, fixation breaks and responses to the incorrect stimuli. Similarly, in **Chapter 3** performance levels plateaued around 70-80% correct (Figure 3.1B). It is unclear what uncertainty the animals ascribe to the outcomes following such erroneous choices, but it seems possible that uncertainty in the cued attention task in **Chapter 2** could be more so described as expected uncertainty. This is likely similar for the reversal learning task in **Chapter 3**, whereby occasional errors at plateau level performance cause something akin to expected uncertainty, while the reversals in reward contingencies, and thereby the accumulation of incorrect outcomes, cause unexpected uncertainty and large prediction errors. Support for this is evident in the relatively slow performance adjustments following a reversal in **Chapter 3** (Figure 3.1B), where it becomes evident that single error occurrences do not usually lead to changes in strategy. Although expected and unexpected uncertainty should promote learning, this process is slower and subtler for expected uncertainty (Yu and Dayan, 2005). In other words, the greater the expected uncertainty – e.g. risk – the lower the learning rate should be, because outcomes are more variable and therefore the resulting prediction errors are less reliable or precise (Preuschoff and Bossaerts, 2007; Bossaerts, 2010). This also means that prediction errors should be scaled according to their reliability or precision, often called 'adaptive' or 'predictive' coding (Friston, 2005; Bossaerts, 2010; Feldman and Friston, 2010), and it has been suggested that attention is the process whereby the brain optimizes these precision estimates (Friston, 2009; Feldman and Friston, 2010). Neurochemically, there is evidence that expected and unexpected uncertainty are supported by the cholinergic and noradrenergic system, respectively (Yu and Dayan, 2003, 2005), and correlates of unexpected uncertainty have for instance been found in the locus coeruleus of the midbrain (Payzan-LeNestour et al., 2013). We find indirect evidence for a dissociation between expected and unexpected uncertainty in our data from **Chapter 4**. Here we find that the FRN, potentially an index of prediction error encoding (Chase et al., 2010; Walsh and Anderson, 2011), is greater during periods of

unexpected uncertainty following a value reversal than during periods in which only expected uncertainty occurs (Figure 4.6). In addition, it has recently been suggested that scaling of prediction errors according to their reliability takes place in dopamine neurons of the midbrain and the striatum (Friston et al., 2012; Diederen and Schultz, 2015; Diederen et al., 2017).

If a main purpose of attention is to assess and reduce uncertainty (Friston, 2009), we find evidence for different mechanisms, likely working in tandem, by which the brain potentially reduces uncertainty via attention at the time of choice/action and during active learning. At the time of choice, during periods of comparatively low uncertainty (e.g. expected uncertainty only), attention towards distracting stimuli is actively reduced over visual cortex (Figure 4.4), which could be supported by enhanced spiketrain correlations that carry information about the target location between ACC and PFC (Figure 2.5). During learning, attention becomes selectively enhanced following low-value outcomes that suggest the need for a behavioral update (e.g. unexpected uncertainty) independent of choice (Figure 4.5), and this is potentially driven by prediction errors computed in frontal (Figure 4.6, Figure 3.6, 3.7) and striatal regions (Figure 3.6, 3.7) that carry information about the valence of the outcome and the specific color value that was selected in the preceding choice.

5.4 Neural network of attention and learning

In order to identify selective contributions to attention and learning, we recorded neural activity in several brain regions that had previously been shown to encode some correlates of attention and/or learning in **Chapters 2** and **3**. This approach allows the direct comparison of neural activity in various brain regions during the same task performance, and it allows the investigation of interactions of neural activity at the single neuron or population level. We find that successful task performance in a volatile environment that employs multiple stimulus features is seemingly supported by a wide array of brain regions that encodes reward relevant and irrelevant information (Figures 3.2, 3.6, 3.7) to potentially support learning and thereby more accurate/efficient attentional deployment (Figure 4.4). Indeed, in **Chapter 3** we find that

more than 50% of neurons in ACC, PFC, CD and VS multiplex information about current or previous chosen task variables at the time of outcome (Figures 3.2, 3.3) whereby variables most likely co-encoded at the second order reflected learning progress and performance in previous trials (Figure 3.3). This culminated in valence-informative and valence-uninformative signals that co-encoded information about the violation of expectation of the current outcome with the specific stimulus feature selected in the preceding choice (Figures 3.6, 3.7). Even though valence-informative prediction error signals linked to the reward-relevant color dimension were most prevalent in ACC, CD and VS, many of the positive prediction error signals were linked to reward-irrelevant features, especially in PFC (Figure 3.7), and valence-uninformative surprise signals were equally likely to carry information about reward – relevant and – irrelevant feature choices (Figure 3.7). It is possible that reward-relevant and reward-irrelevant prediction error signals are differentially integrated in ACC and PFC, potentially at the single neuron level (Donahue and Lee, 2015), or through spiketrain correlations between the two areas (Figure 2.2). Although speculative, this could mean that information about violations of expectation, and task-relevant and -irrelevant feature choices are orchestrated at the level of the frontal cortex, especially in ACC, where they become translated into adjustments in attention to allow for flexible goal-directed behavior (Hyman et al., 2017; Leong et al., 2017). Leong and colleagues (2017) recently found evidence for the dynamic interaction between prediction errors and attention using fMRI in humans, whereby prediction errors were biased by attention, and attention in fronto-parietal networks changed based on learning. In turn, Hyman et al. (2017) showed that prediction error activity in ACC was directly related to a following switch in behavior in a reversal learning paradigm in rats.

5.5 Interpretational limitations

It is not without challenge to compare the results from the two single unit studies in monkeys in **Chapters 2** and **3** with the human event-related potential (ERP) results from **Chapter 4**. The EEG signal is thought to represent the extracellular currents reflecting the summed dendritic postsynaptic potentials of millions

of parallel-aligned pyramidal cells (Buzsáki et al., 2012; Lopes da Silva, 2013). It is therefore almost impossible to generalize from the level of single neuron activity to the EEG signal observed at the skull. In addition, the content of the EEG, such as the functional and anatomical configurations of the underlying micro-circuits, is unknown (Cohen, 2017). This means it is unknown what specific circuit-level neuronal activity underlies, for instance, the N2pc component observed over visual cortex. This in turn means that any interpretation that bridges activity of single neurons with activity observed in the EEG, such as the possibility that frontal feature-specific prediction error signals observed in the macaque may underlie the amplitude of the FRN observed in humans in a similar task, is speculative at best. The spiking activity of single neurons represents almost the opposite extreme of the electrophysiological spectrum. Therefore, their interpretation is also limited when interactions at the circuit level are not considered, and one needs to be careful with inferring functionality at the level of an entire anatomical area when it is unclear how signals from single neurons are integrated at the population level. This means for instance that the feature-specific prediction error signals observed in ACC, PFC, CD and VS in **Chapter 3** may be differently integrated functionally depending on the area, and may therefore likely serve somewhat different functions in the different areas.

5.6 Future directions

There are three lines of inquiries that immediately result from the research described in this thesis. 1) In order to better understand the functional role of inter-areal spiketrain correlations in fronto-cingulate cortex, it would be valuable and highly informative to know how spiketrain correlations develop with learning and how they correlate with performance. I speculate that spiketrain correlations between ACC and PFC increase in magnitude with successful learning of the attentional target and that they are reduced on error trials. 2) In order to understand the functional integration of feature-specific prediction errors at the circuit level in frontal and striatal regions, it would be important for advancing theoretical accounts to test the spike-field coherence of those neurons encoding feature-specific prediction error signals. In

particular, it would be important to know whether spike-field coherence differs for surprise signals that are uninformative about the valence of the outcome, but potentially serve to increase attention in the following trials, compared with positive and negative prediction error signals that are informative about outcome valence and the relevant feature selected in the preceding choice, that may serve to identify those synapses that need updating during learning. In addition, it has not yet been tested whether and how LFP coherence between frontal and striatal sub-regions changes in trials following large versus small prediction error signals. It would be important to see whether coherence increases as a function of prediction error size to potentially support flexible behavioral updating. 3) To further understand the underlying neuronal mechanisms of flexible reversal learning in humans, it would be crucial to test whether long-range oscillations, for instance in the gamma or beta frequency band (Siegel et al., 2012), between frontal and parietal or visual areas flexibly track learning performance in the volatile reversal learning environment employed here. It could be possible, for instance, that distant gamma coherence in frontal and visual areas changes as a function of uncertainty during periods of active behavioral evaluation and adjustment.

Insights into the above aspects could help to discern the circuit and network level mechanisms of flexible adjustments of behavior during learning. The cell type analyses in **Chapters 2** and **3** for instance point to cell-specific contributions to the network level implementation of behavioral flexibility.

5.7 Implications for understanding brain health

Delineating the neural mechanisms that underlie attention and learning in the healthy brain could be critical to identifying the sources of failure in functions of attention and learning in various major psychiatric disorders, such as depression, schizophrenia and psychosis. These disorders are characterized by dysfunctions in, for instance, the cortico-striatal networks, leading to severe disturbances in attentional control and valuation processes (Price and Drevets, 2012). Patients with major depressive disorder fail to

develop a response bias towards rewarding stimuli, and show an oversensitivity to negative feedback in probabilistic reward tasks, with a tendency to bias processing of stimuli with negative meaning compared with positive or neutral meaning (Henriques et al., 1994; Elliott et al., 2000; Tavares et al., 2008). In line with this, regions of the ACC, PFC, and ventral striatum show reduced BOLD activity during reward learning, probabilistic reversal learning tasks, and delivery of unexpected reward, in depressed patients (Tavares et al., 2008; Wacker et al., 2009; Robinson et al., 2012). A potential mechanism that is thought to underlie psychosis – a symptom of e.g. schizophrenia, is aberrant processing of prediction error signals, especially in the sensory domain (Corlett et al., 2011). Early symptoms of psychosis for instance include increased perceptual experiences, such as sounds that appear louder and colors that appear brighter, consistent with abnormally large prediction errors, while delusions may result when an individual experiences abnormally large prediction errors for events that should be unsurprising and highly predictable (Corlett et al., 2011). Negative symptoms of schizophrenia that involve deficits in motivated behavior may be similarly grounded – patients with negative symptoms seem to have problems predicting consequences of series of actions (Polgár et al., 2008) and using those predictions to guide goal-directed behavior (Barch et al., 2003). In line with this, schizophrenic patients show decreased activity in regions of the PFC during flexible behavioral tasks and aberrant synthesis and transmission of dopaminergic function in the striatum (Lindström et al., 1999; Barch et al., 2001; Meyer-Lindenberg et al., 2002). The previous highlights the importance of understanding the neural mechanisms of cognitive functions, such as attention and learning, in the healthy brain to potentially allow to better understand the failure of these systems in the un-healthy brain.

5.8 Concluding remarks

As a whole, this dissertation has provided valuable new insights into some of the neural mechanisms that underlie the successful control and learning of attention. We have studied several brain regions that have been shown to encode, and the neural signatures that have been shown to reflect, aspects of attention and

learning. We have found evidence that when learning is minimal during active control of attention, spiketrain correlations between neurons that prefer contra-lateral attention shifts in ACC and PFC are increased, and the event-related potential N2pc over visual cortex is reduced for distracting stimuli. We also found evidence that when behavior needs to be actively adjusted, prediction error signals in frontal and striatal subregions that carry specific information about the valence of the outcome and the preceding feature choice may aid learning by assigning credit to the source of the unexpected outcome, and attention, measured with the N2pc, is selectively increased following such unexpected outcomes. Attention and learning are highly inter-related processes that seem to involve a wide array of brain regions. In particular **Chapter 3** revealed the similarity across brain regions in encoding relevant information to task performance, suggesting that differences in specific contributions to attention and learning are potentially subtler than anticipated. Likely driven by the dense interconnectivity between these brain regions (Haber and Knutson, 2010), this highlights the complexity of network interactions between frontal and striatal areas during attention and learning, warranting further investigations into the specific interactions between these brain regions during learning and control of attention. Indeed, we must likely study a host of additional brain regions, such as the amygdala (Averbeck and Costa, 2017), the basal forebrain (Monosov et al., 2015), and hippocampus (Long et al., 2016), and investigate in more detail the specific role of neuromodulators (Wilson et al., 2016), to fully elucidate how we determine what to attend to in our environment.

5.9 References

Averbeck BB, Costa VD (2017) Motivational neural circuits underlying reinforcement learning. Nat Neurosci 20:505–512.

Barch DM, Carter CS, Braver TS, Sabb FW, MacDonald III A, Noll DC, Cohen JD (2001) Selective deficits in prefrontal cortex function in medication-naive patients with schizophrenia. Arch Gen Psychiatry 58:280–288.

Barch DM, Carter CS, MacDonald AW, Braver TS, Cohen JD (2003) Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. J Abnorm Psychol 112:132–143.

Bossaerts P (2010) Risk and risk prediction error signals in anterior insula. Brain Struct Funct:1–9.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents - EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Chase HW, Swainson R, Durham L, Benham L, Cools R (2010) Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. J Cogn Neurosci 23:936–946.

Cohen MX (2017) Where Does EEG Come From and What Does It Mean? Trends Neurosci 40:208–218.

Corlett PR, Honey GD, Krystal JH, Fletcher PC (2011) Glutamatergic model psychoses: prediction error, learning, and inference. Neuropsychopharmacology 36:294–315.

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cogn Affect Behav Neurosci 8:429–453.

Diederen KMJ, Schultz W (2015) Scaling prediction errors to reward variability benefits error-driven learning in humans. J Neurophysiol 114:1628–1640.

Diederen KMJ, Ziauddeen H, Vestergaard MD, Spencer T, Schultz W, Fletcher PC (2017) Dopamine Modulates Adaptive Prediction Error Coding in the Human Midbrain and Striatum. J Neurosci 37:1708–1720.

Donahue CH, Lee D (2015) Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. Nat Neurosci 18:1–9.

Elliott R, Rubinsztein JS, Sahakian BJ, Dolan RJ (2000) Selective attention to emotional stimuli in a verbal go/no-go task: an fMRI study. Neuroreport 11:1739–1744.

Feldman H, Friston KJ (2010) Attention, Uncertainty, and Free-Energy. Front Hum Neurosci 4:1–23.

Friston KJ (2005) A theory of cortical responses. Philos Trans R Soc B 360:815–836.

Friston KJ (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293–301.

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S (2012) Dopamine, affordance and active inference. PLoS Comput Biol 8.

Gottlieb J (2012) Attention, learning and the value of information. Neuron 76:281–295.

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26.

Henriques JB, Glowacki JM, Davidson RJ (1994) Reward fails to alter response bias in depression. J Abnorm Psychol 103:460–466.

Hyman JM, Holroyd CB, Seamans JK (2017) A Novel Neural Prediction Error Found in Anterior Cingulate Cortex Ensembles. Neuron 95:447–456.e3.

Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y (2017) Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron 93:451–463.

Lindström LH, Gefvert O, Hagberg G, Lundberg T, Bergström M, Hartvig P, Långström B (1999) Increased dopamine synthesis rate in medial prefrontal cortex and striatum in schizophrenia indicated by L-(β-11C) DOPA and PET. Biol Psychiatry 46:681–688.

Long NM, Lee H, Kuhl BA (2016) Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. J Neurosci 36:1850–16.

Lopes da Silva F (2013) EEG and MEG: Relevance to neuroscience. Neuron 80:1112–1128.

Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev 82:276–298.

Meyer-Lindenberg A, Miletich RS, Kohn PD, Esposito G, Carson RE, Quarantelli M, Weinberger DR, Berman KF (2002) Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. Nat Neurosci 5:267–271.

Miltner WHR, Braun CH, Coles MGH (1997) Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a "Generic" Neural System for Error Detection. J Cogn Neurosci 9:788–798.

Monosov IE, Leopold DA, Hikosaka O (2015) Neurons in the Primate Medial Basal Forebrain Signal Combined Information about Reward Uncertainty, Value, and Punishment Anticipation. J Neurosci 35:7443–7459

Payzan-Lenestour E, Bossaerts P (2011) Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput Biol 7.

Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP (2013) The neural representation of unexpected uncertainty during value-based decision making. Neuron 79:191–201.

Polgár P, Farkas M, Nagy O, Kelemen O, Réthelyi J, Bitter I, Myers CE, Gluck MA, Kéri S (2008) How to find the way out from four rooms? The learning of "chaining" associations may shed light on the neuropsychology of the deficit syndrome of schizophrenia. Schizophr Res 99:200–207.

Preuschoff K, Bossaerts P (2007) Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci 1104:135–146.

Price JL, Drevets WC (2012) Neural circuits underlying the pathophysiology of mood disorders. Trends Cogn Sci 16:61–71.

Robinson OJ, Cools R, Carlisi CO, Sahakian BJ, Drevets WC (2012) Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. Am J Psychiatry 169:152–159.

Siegel M, Donner TH, Engel AK (2012) Spectral fingerprints of large-scale neuronal interactions. Nat Rev Neurosci 13:121–134.

Tavares JVT, Clark L, Furey ML, Williams GB, Sahakian BJ, Drevets WC (2008) Neural basis of abnormal response to negative feedback in unmedicated mood disorders. Neuroimage 42:1118–1126.

Wacker J, Dillon DG, Pizzagalli DA (2009) The role of the nucleus accumbens and rostral anterior cingulate cortex in anhedonia: Integration of resting EEG, fMRI, and volumetric techniques.

Neuroimage 46:327–337.

Walsh MM, Anderson JR (2011) Modulation of the feedback-related negativity by instruction and experience. Proc Natl Acad Sci USA 208:19048–19035.

Wilson CRE, Vezoli J, Stoll FM, Faraut MCM, Leviel V, Knoblauch K, Procyk E (2016) Prefrontal Markers and Cognitive Performance Are Dissociated during Progressive Dopamine Lesion. PLoS Biol 14:1–31.

Yu AJ, Dayan P (2003) Expected and unexpected uncertainty: ACh and NE in the neocortex. Adv Neural Inf Process Syst:173–180.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.

Appendix A

Supplementary Figures Chapter 3

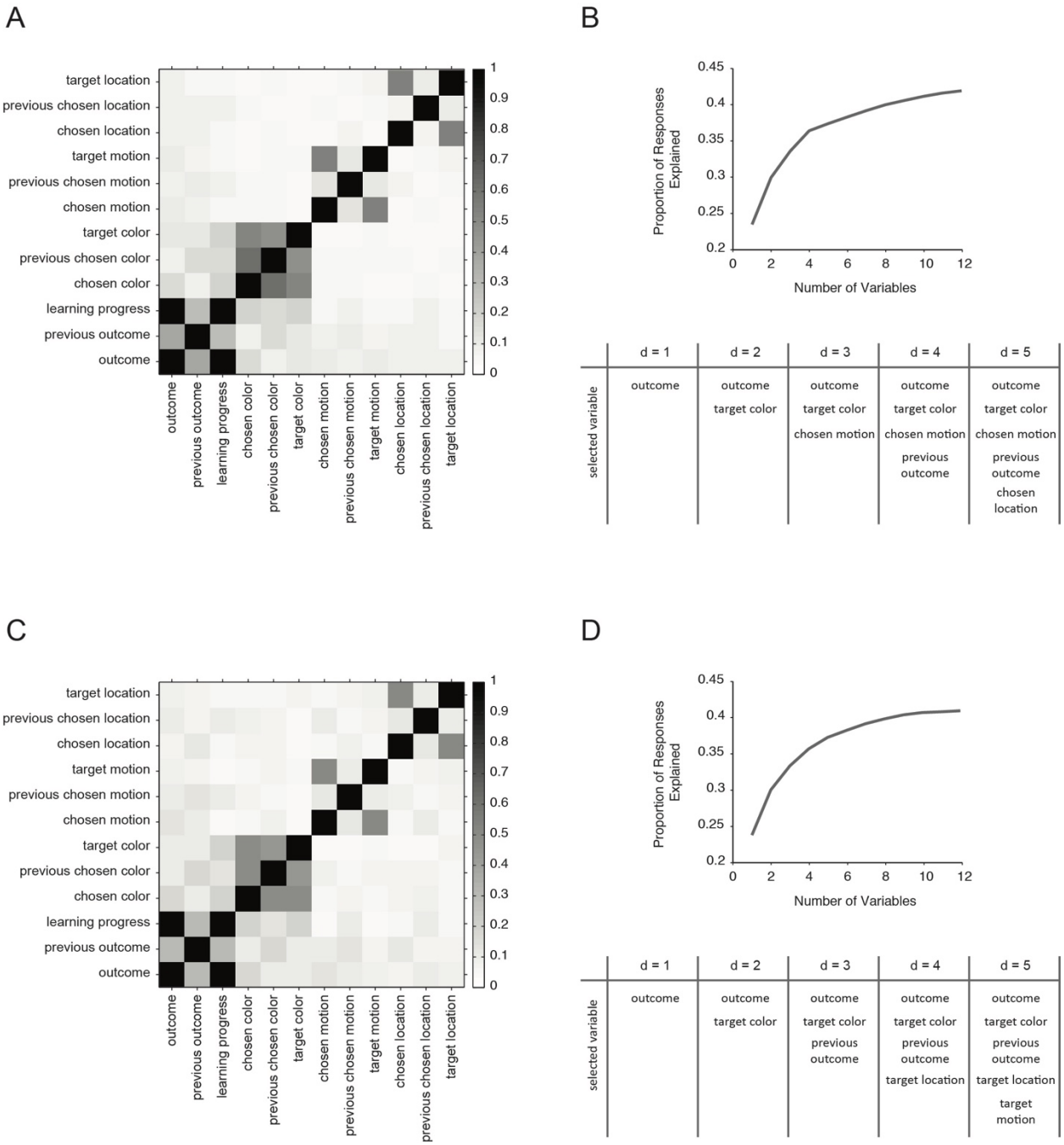

Figure S 1 Variable correlation and variable best-subset selection.

A Correlation matrix for all 12 variables tested. Elements of the correlation matrix vary between 0 and 1. Note that the variables outcome and learning progress are shown with a correlation of 1, because the variable learning progress contained only correct trials. **B** *Top:* Proportion of responses explained for each added variable based on the best-subset selection procedure (see Methods). *Bottom:* Best variable subsets for the first 5 variable combinations that explain the largest number of responses. **C** and **D** as **A** and **B** for *monkey K*.

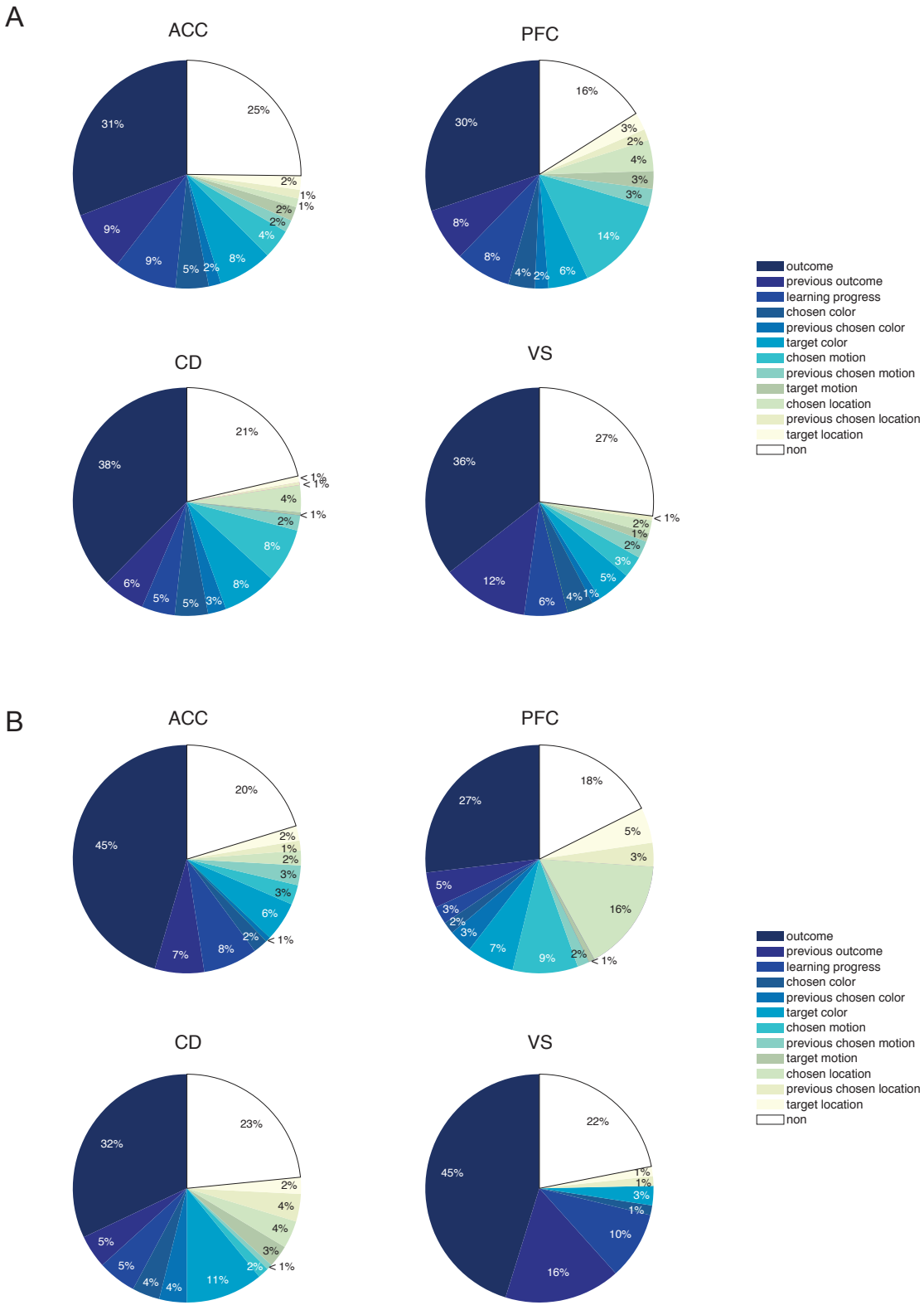

Figure S 2 Task variables encoded split by areas.

A Proportion of task variables encoded at the time of outcome for neurons in ACC, PFC, CD, and VS for *monkey H*. **B** Same as **A** for *monkey K*.

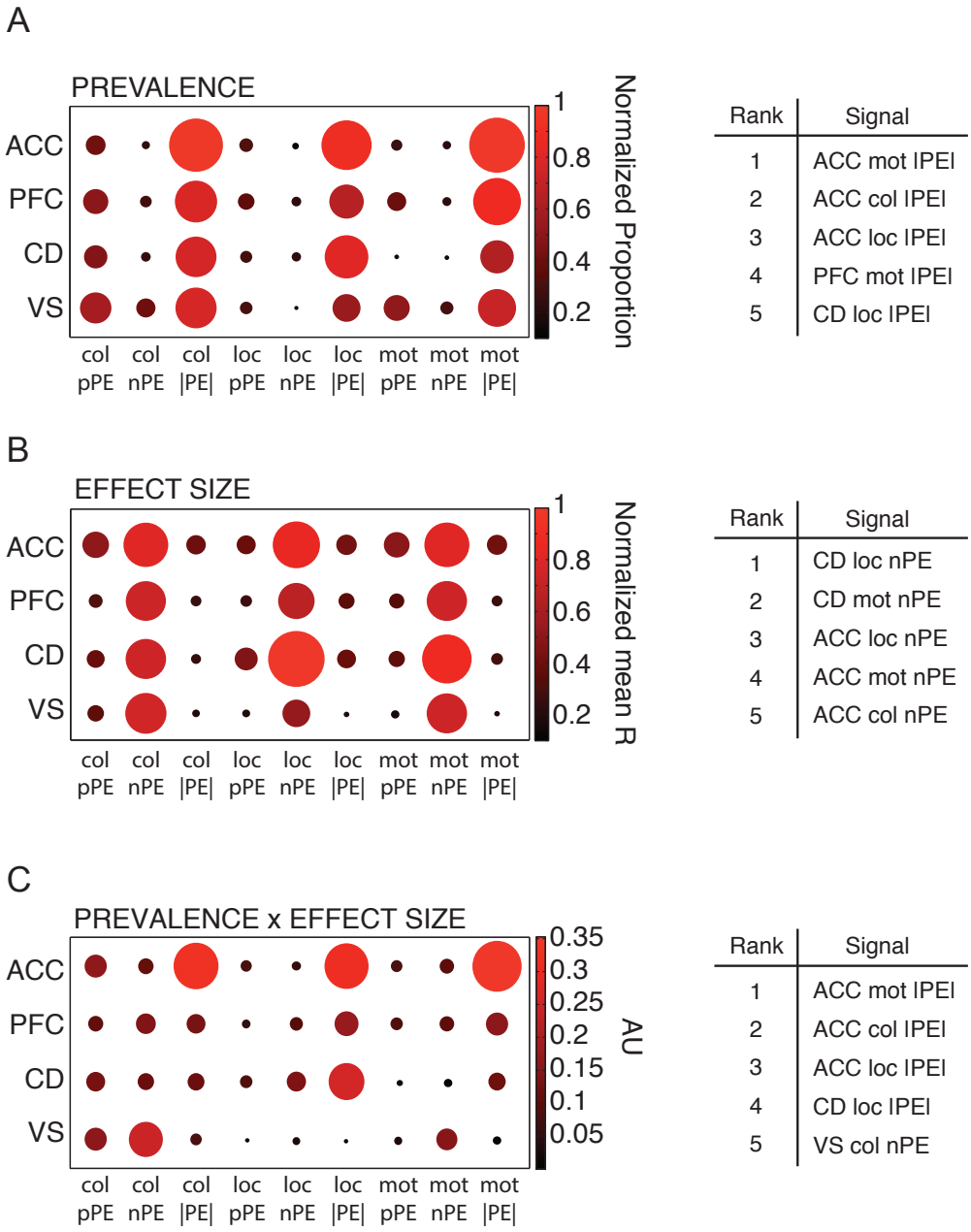

Figure S 3 Prevalence and effect sizes of feature-specific PE signals.

A *Left:* Normalized proportions of feature-specific PE signals illustrated by circle size as well as color. Proportions of PE signals were normalized between all areas, PE types and features to scale between 0.1 and 1. *Right:* First five rank-ordered feature-specific PE signals according to their prevalence. **B** *Left:* Normalized effect size of feature-specific PE signals (mean Spearman's R) illustrated by circle size as well as color. Effect sizes of PE signals were normalized between all areas, PE types and features to scale

between 0.1 and 1. *Right:* First five rank-ordered feature-specific PE signals according to their effect size. **C** *Left:* Product of normalized prevalence (A) and normalized effect size (B). *Right:* First five rank-ordered feature-specific PE signals according to their value of [prevalence x effect size]. A-C are shown for data combined across *monkey H* and *K*.

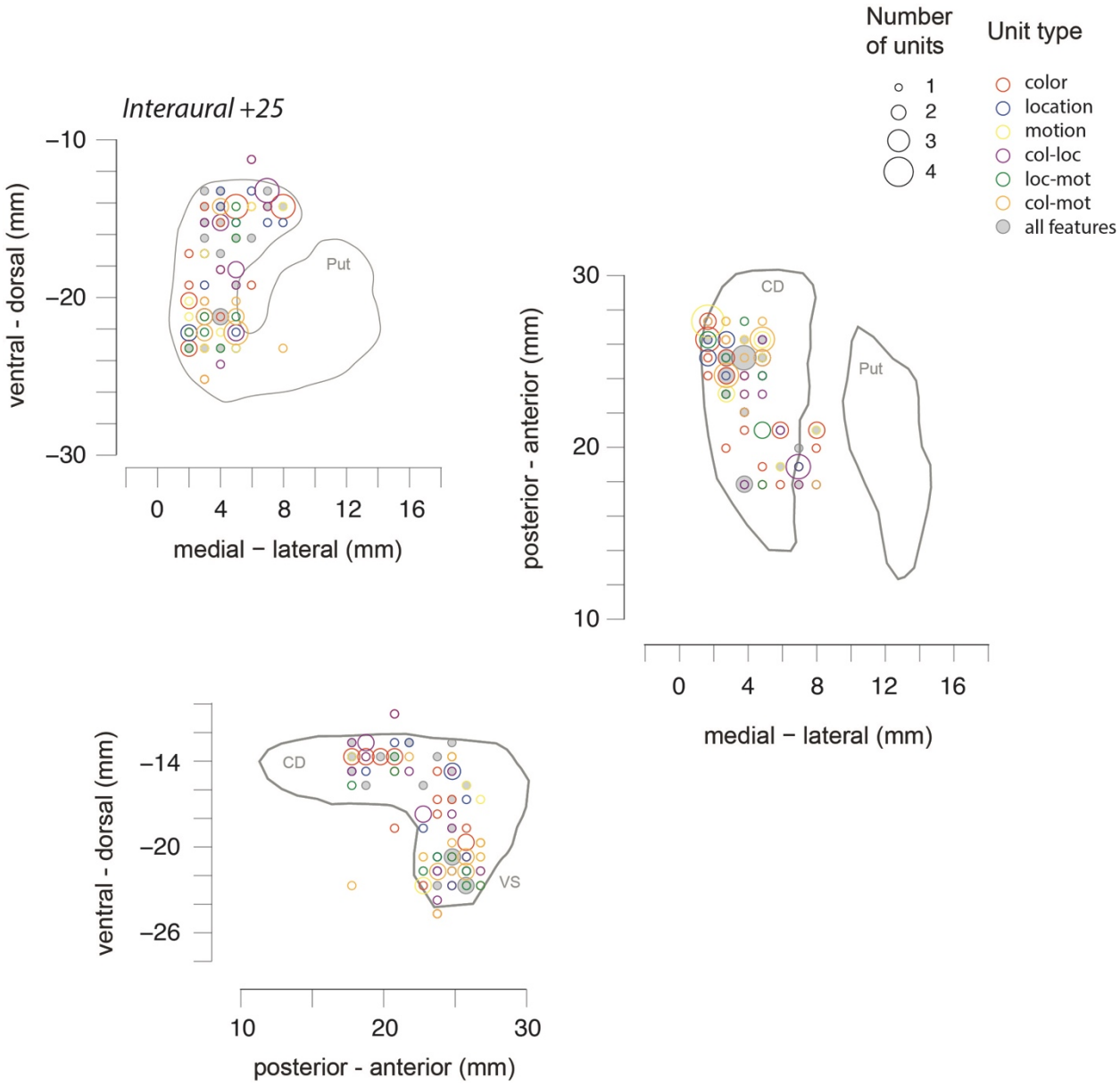

Figure S 4 Anatomical locations in striatum of feature-specific PEs in *monkey H*.

Feature-specific PE encoding neurons are differentiated with colors based on the feature-type PE they encoded and whether they encoded more than one feature-type PE. Displayed are three orientations of the caudate and ventral striatum: coronal view (top left), axial view (right), and sagittal view (bottom left). Within a given orientation, units are collapsed across the third non-visible orientation. Coordinates are relative to stereo-tactic zero.

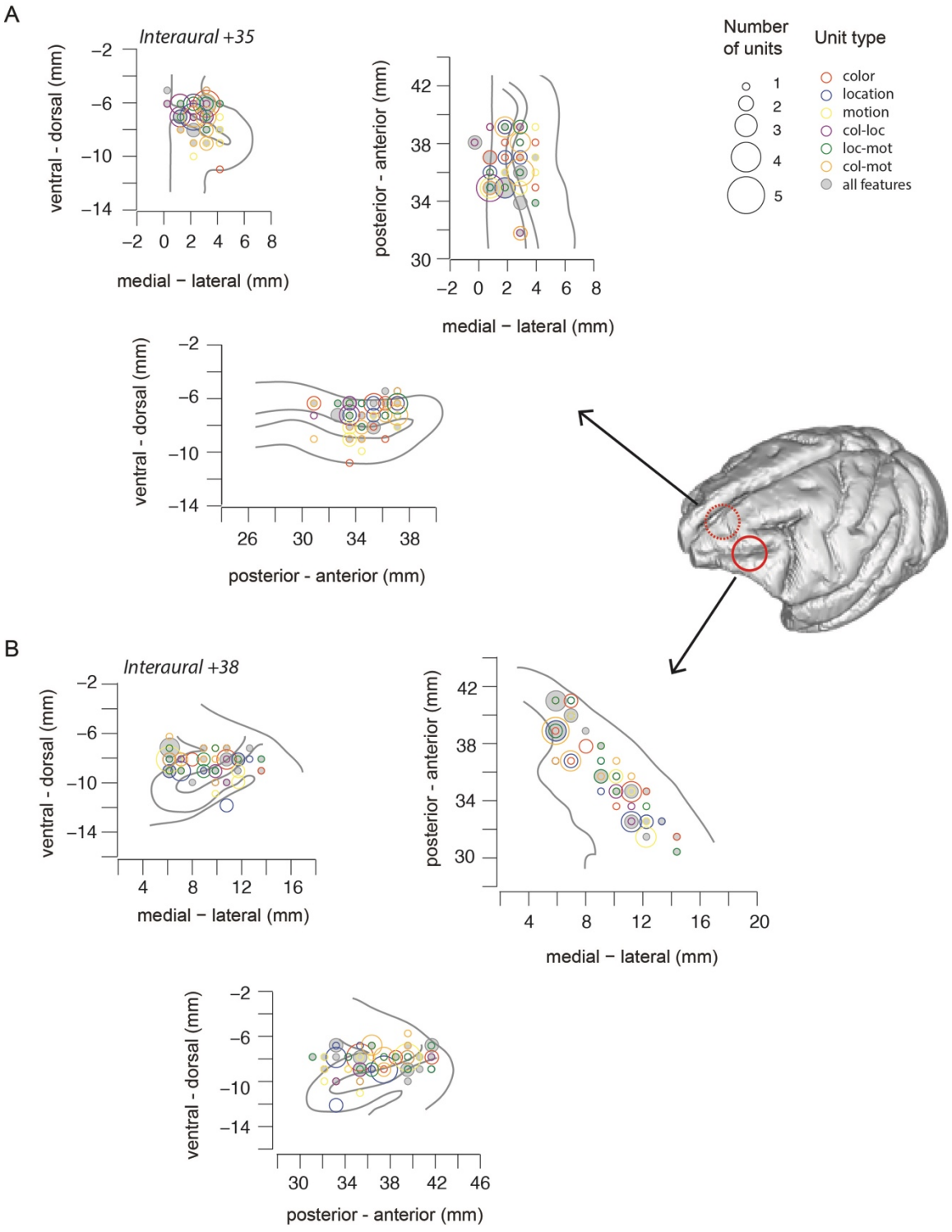

Figure S 5 Anatomical locations in ACC and PFC of feature-specific PEs in *monkey H*.

Same conventions as in Figure S4 for ACC locations (**A**) and PFC locations (**B**).

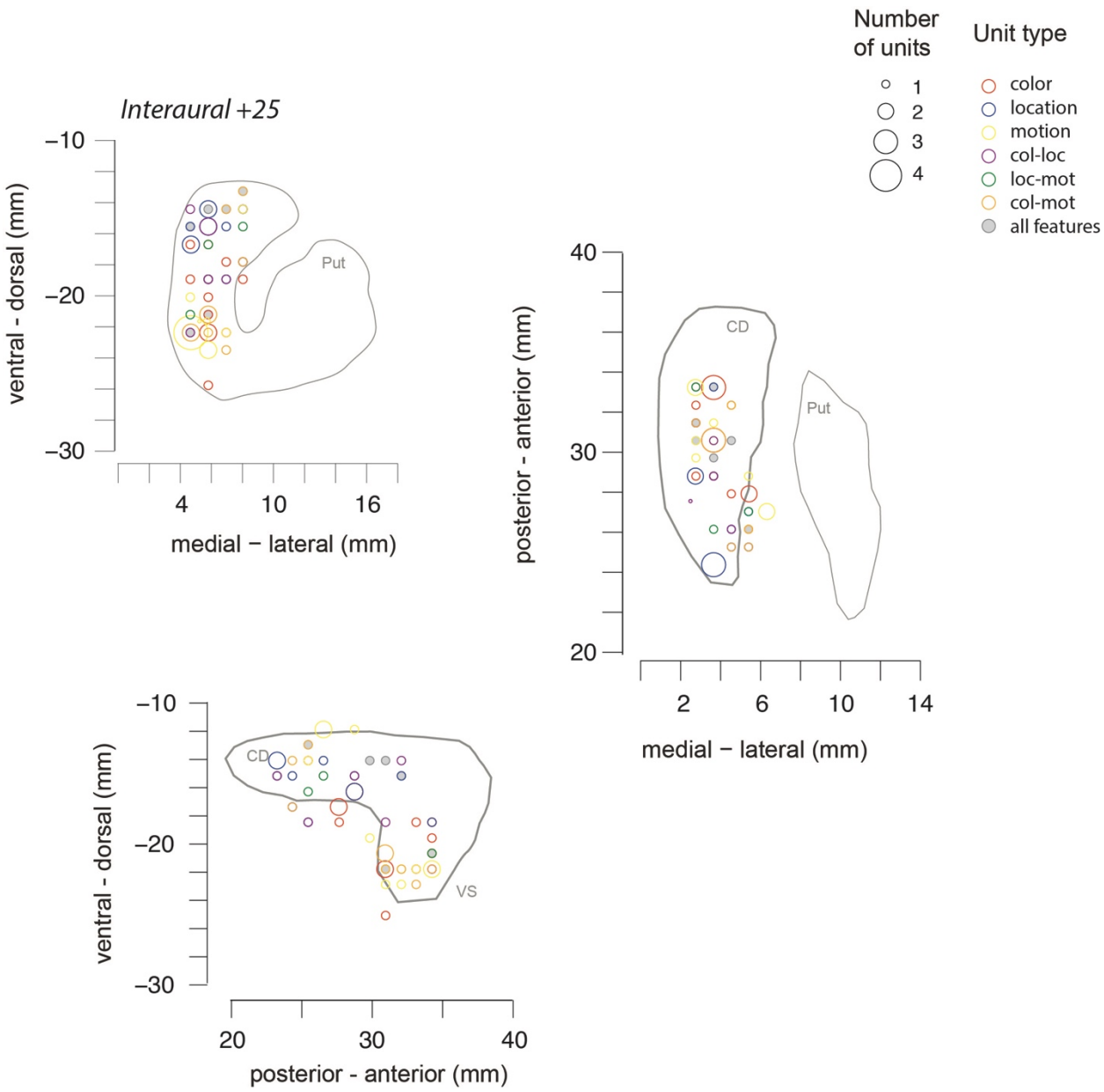

Figure S 6 Anatomical locations in striatum of feature-specific PEs in *monkey K*.

Same conventions as in Figure S4.

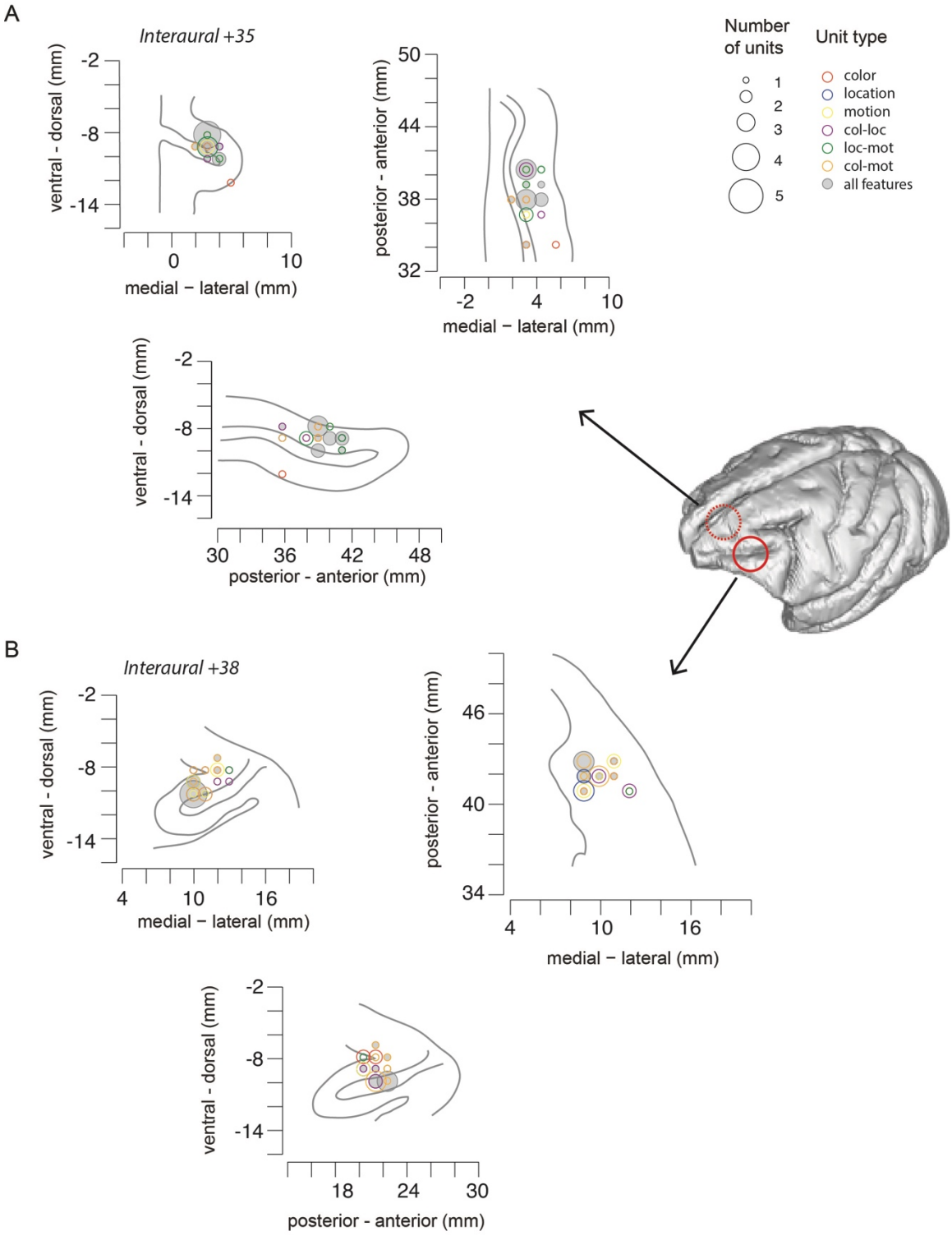

Figure S 7 Anatomical locations in ACC and PFC of feature-specific PEs in *monkey K*.

Same conventions as in Figure S5.

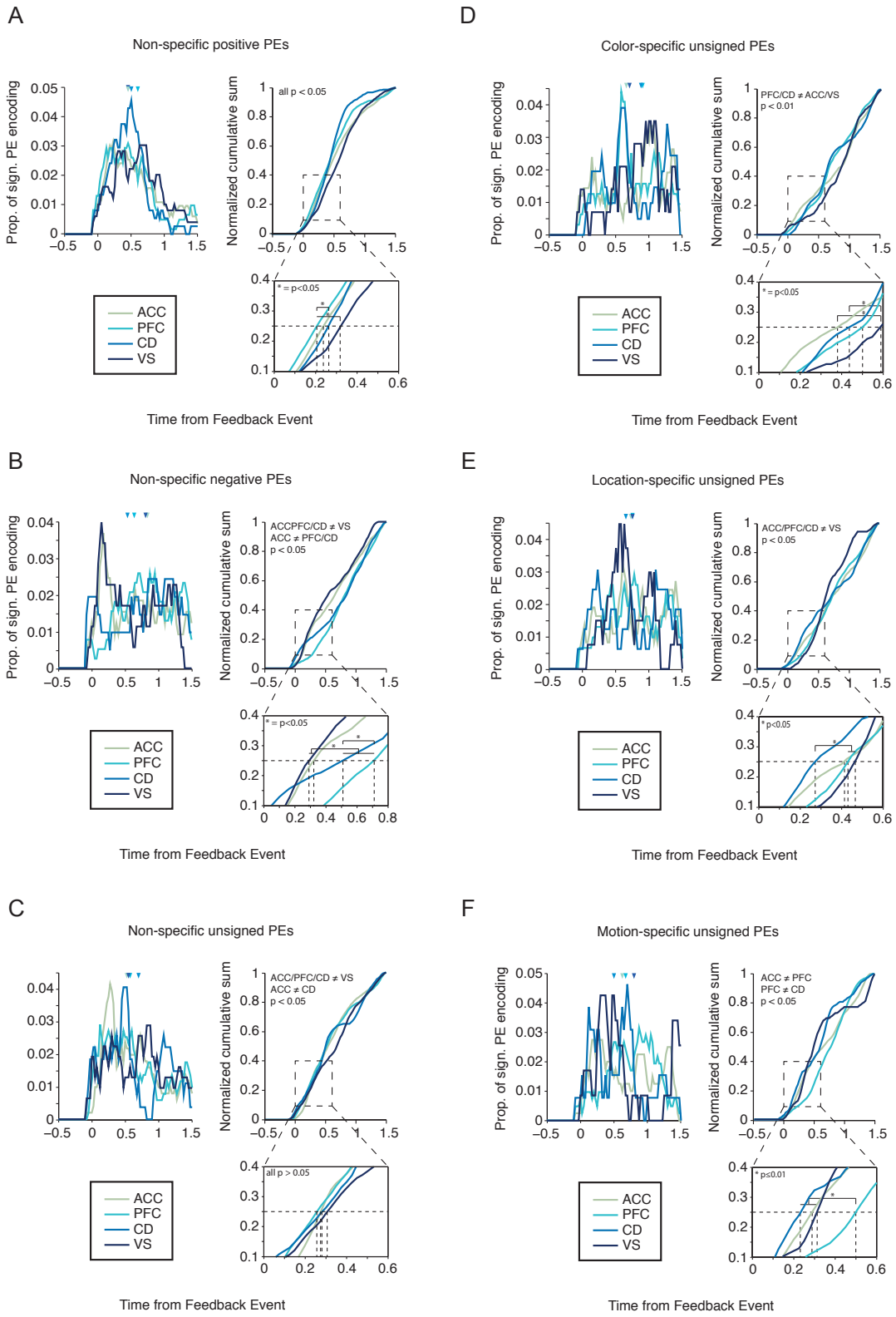

Figure S 8 Latencies of non-specific PEs and feature-specific unsigned PEs in *monkey H*.

Conventions are equivalent to Figure 3.9. Latency analyses are shown for non-specific positive PEs (A), non-specific negative PEs (B), and non-specific unsigned PEs (C). Latency analyses are also shown for color-specific unsigned PEs (D), location-specific unsigned PEs (E), and motion-specific unsigned PEs (F).

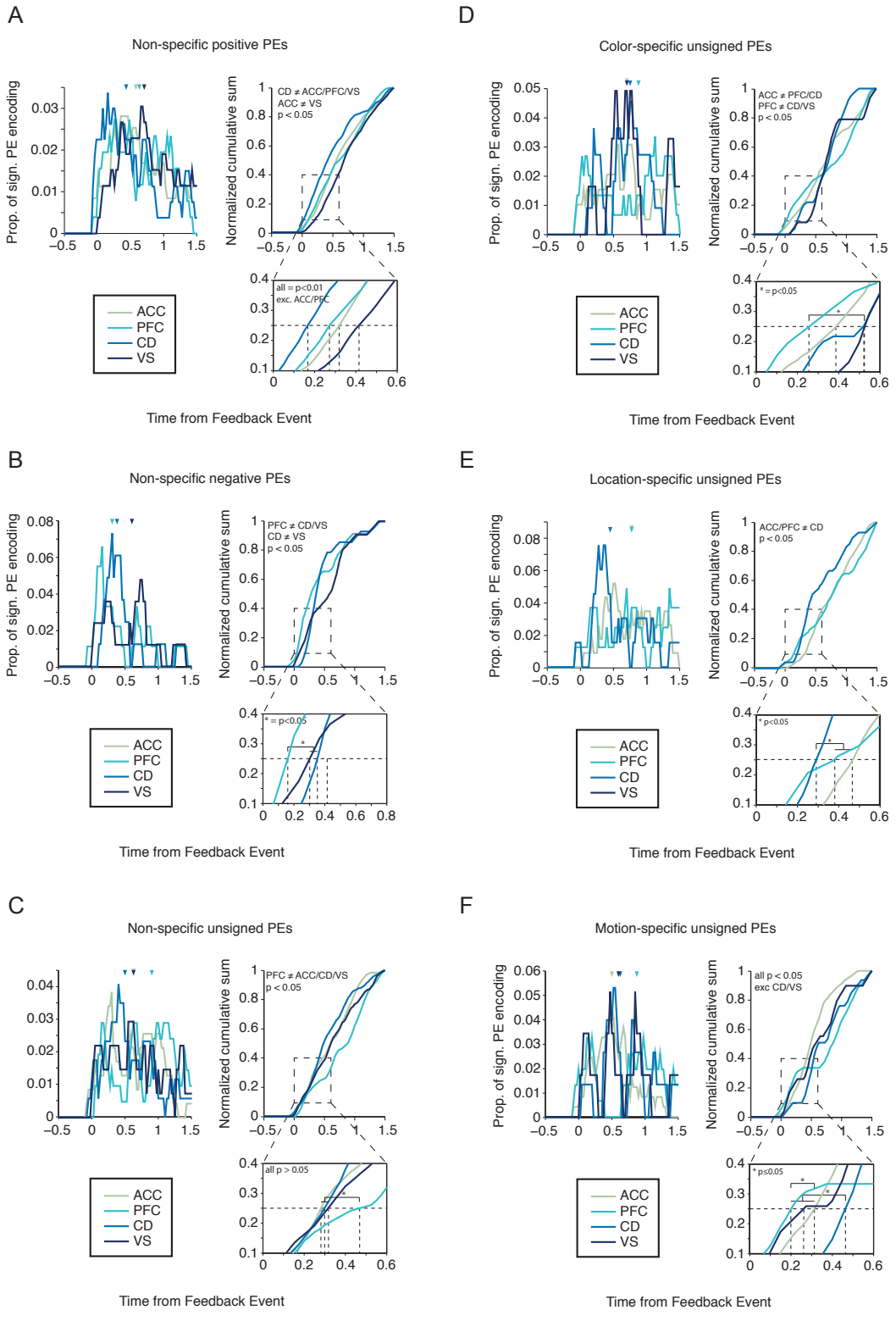

Figure S 9 Latencies of non-specific PEs and feature-specific unsigned PEs in *monkey K*.

Same conventions as in Figure S7.

Appendix B
Supplementary Methods and Results Chapter 3

Supplementary Methods

Stepwise selection method. We determined for each neuron the variable that provided the best-fit (highest R^2 value) and then selected the variable that provided the highest number of best fits across neurons. We then removed this variable and all neurons that it explained ($R^2 < 0.05$) in order to re-compute the best-fit for the remaining neurons. We continued these iterations until no further variable explained a minimum of 5% of all neural responses. Results obtained with the stepwise selection method may be 'path-dependent', because variables are removed in order and may therefore not yield the best possible subset of variables.

Supplementary Results

Stepwise variable selection

Results obtained with the stepwise selection method were very similar to those obtained with the best-subset method. The stepwise variable selection method revealed that the variables **outcome**, **previous outcome**, and **chosen motion** each explained more than 5% of neural responses for *monkey H*, and the variables **outcome**, **previous outcome**, and **target color** explained more than 5% of neural responses for *monkey K*.

Time courses of prediction error encoding – comparisons of cumulative sums

In *monkey H*, cumulative sums of *color-specific positive* PEs differed between ACC and PFC and between ACC and CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.016$; $p_{(ACC\text{-}VS)} = 0.17$; $p_{(PFC\text{-}CD)} = 0.202$; $p_{(PFC\text{-}VS)} = 0.051$; $p_{(CD\text{-}VS)} = 0.152$) (Figure 3.9A). Cumulative sums for *non-specific positive* PEs differed between all areas (Kolmogorov

Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.019$; $p_{(ACC\text{-}CD)} < 0.001$; $p_{(ACC\text{-}VS)} = 0.015$; $p_{(PFC\text{-}CD)} = 0.019$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} < 0.001$; Figure S7A Appendix A).

Cumulative sums of *color-specific negative* PEs differed between all areas except between PFC and CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.002$; $p_{(ACC\text{-}CD)} = 0.009$; $p_{(ACC\text{-}VS)} < 0.001$; $p_{(PFC\text{-}CD)} = 0.225$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} < 0.001$) (Figure 3.9B). Cumulative sums of *non-specific negative* PEs also differed between all areas with the exception of ACC-VS and PFC-CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.042$; $p_{(ACC\text{-}VS)} = 0.415$; $p_{(PFC\text{-}CD)} = 0.054$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} = 0.003$) (Figure S7B Appendix A).

Cumulative sums of *feature-specific unsigned* PEs differed between ACC-PFC, ACC-CD and PFC-CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.001$; $p_{(ACC\text{-}VS)} = 0.080$; $p_{(PFC\text{-}CD)} = 0.002$; $p_{(PFC\text{-}VS)} = 0.069$; $p_{(CD\text{-}VS)} = 0.080$) (Figure 3.9C). Cumulative sums in *non-specific unsigned* PEs also differed between all areas except for ACC-PFC and PFC-CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.164$; $p_{(ACC\text{-}CD)} = 0.032$; $p_{(ACC\text{-}VS)} = 0.032$; $p_{(PFC\text{-}CD)} = 0.067$; $p_{(PFC\text{-}VS)} = 0.001$; $p_{(CD\text{-}VS)} = 0.005$) (Figure S7C Appendix A).

Cumulative sums of *color-specific unsigned* PEs differed between all areas except for ACC-VS and PFC-CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.009$; $p_{(ACC\text{-}CD)} = 0.009$; $p_{(ACC\text{-}VS)} = 0.146$; $p_{(PFC\text{-}CD)} = 0.146$; $p_{(PFC\text{-}VS)} = 0.008$; $p_{(CD\text{-}VS)} = 0.003$) (Figure S7D Appendix A). Cumulative sums of *location-specific unsigned* PEs differed between VS and all other areas (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.429$; $p_{(ACC\text{-}CD)} = 0.361$; $p_{(ACC\text{-}VS)} = 0.001$; $p_{(PFC\text{-}CD)} = 0.146$; $p_{(PFC\text{-}VS)} = 0.037$; $p_{(CD\text{-}VS)} = 0.002$) (Figure S7E Appendix A). Cumulative sums of *motion-specific unsigned* PEs differed between all areas except for CD-VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.002$; $p_{(ACC\text{-}VS)} = 0.011$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} = 0.111$) (Figure S7F Appendix A).

In *monkey K*, cumulative sums of *color-specific positive* PEs differed between all areas except ACC- PFC (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.172$; $p_{(ACC\text{-}CD)} < 0.001$; $p_{(ACC\text{-}VS)} = 0.006$; $p_{(PFC\text{-}CD)} = 0.001$; $p_{(PFC\text{-}VS)} = 0.031$; $p_{(CD\text{-}VS)} < 0.001$) (Figure 3.9D). Cumulative sums for *non-specific positive* PEs also differed between all areas except ACC-PFC and PFC-VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.498$; $p_{(ACC\text{-}CD)} = 0.003$; $p_{(ACC\text{-}VS)} = 0.028$; $p_{(PFC\text{-}CD)} = 0.003$; $p_{(PFC\text{-}VS)} = 0.050$; $p_{(CD\text{-}VS)} < 0.001$; Figure S8A Appendix A).

Cumulative sums of *color-specific negative* PEs differed between all areas except between ACC-VS and CD-VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.004$; $p_{(ACC\text{-}CD)} = 0.004$; $p_{(ACC\text{-}VS)} = 0.135$; $p_{(PFC\text{-}CD)} = 0.018$; $p_{(PFC\text{-}VS)} = 0.004$; $p_{(CD\text{-}VS)} = 0.135$) (Figure 3.9E). Cumulative sums of *non-specific negative* PEs also differed between PFC and CD and between CD and VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.060$; $p_{(ACC\text{-}CD)} = 0.075$; $p_{(ACC\text{-}VS)} = 0.704$; $p_{(PFC\text{-}CD)} = 0.009$; $p_{(PFC\text{-}VS)} = 0.074$; $p_{(CD\text{-}VS)} = 0.002$) (Figure S8B Appendix A).

Cumulative sums of *feature-specific unsigned* PEs differed between ACC-PFC and between PFC-CD (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.399$; $p_{(ACC\text{-}VS)} = 0.228$; $p_{(PFC\text{-}CD)} <= 0.001$; $p_{(PFC\text{-}VS)} = 0.075$; $p_{(CD\text{-}VS)} = 0.399$) (Figure 3.9F). Cumulative sums in *non-specific unsigned* PEs differed between ACC and PFC, between PFC and CD and between PFC and VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.347$; $p_{(ACC\text{-}VS)} = 0.327$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} = 0.010$; $p_{(CD\text{-}VS)} = 0.347$) (Figure S8C Appendix A).

Cumulative sums of *feature-specific unsigned* PEs differed between all areas except for ACC-VS and CD-VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.024$; $p_{(ACC\text{-}CD)} = 0.023$; $p_{(ACC\text{-}VS)} = 0.117$; $p_{(PFC\text{-}CD)} < 0.001$; $p_{(PFC\text{-}VS)} = 0.019$; $p_{(CD\text{-}VS)} = 0.097$) (Figure S8D). Cumulative sums of *location-specific unsigned* PEs differed between all areas except ACC-PFC (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} = 0.082$; $p_{(ACC\text{-}CD)}$

< 0.001; $p_{(ACC\text{-}VS)} < 0.001$; $p_{(PFC\text{-}CD)} = 0.001$; $p_{(PFC\text{-}VS)} < 0.001$; $p_{(CD\text{-}VS)} < 0.001$) (Figure S8E Appendix A). Cumulative sums of *motion-specific unsigned* PEs differed between all areas except for CD-VS (Kolmogorov Smirnoff test, Bonferroni-Holm multiple comparison correction: $p_{(ACC\text{-}PFC)} < 0.001$; $p_{(ACC\text{-}CD)} = 0.001$; $p_{(ACC\text{-}VS)} = 0.049$; $p_{(PFC\text{-}CD)} = 0.027$; $p_{(PFC\text{-}VS)} = 0.049$; $p_{(CD\text{-}VS)} = 0.306$) (Figure S8F Appendix A).

Appendix C

Co-Authorship

The research for **Chapter 2** was designed by Dr. Thilo Womelsdorf, and all data collection was done at Western University, London, ON by Dr. Thilo Womelsdorf. Analyses were performed by me with constructive feedback and help from Dr. Thilo Womelsdorf and Dr. Stephanie Westendorff. The manuscript was drafted by me and edited by all authors.

The research for **Chapter 3** was designed by Dr. Thilo Womelsdorf and Dr. Stephanie Westendorff. Animal training and data collection were performed by me (with help from Dr. Stephanie Westendorff, Samira Azimi, and Ali Hassani). Reinforcement learning modelling was done by Dr. Salva Ardid, Dr. Paul Tiesinga, and Dr. Thilo Womelsdorf. All other analyses were performed by me with constructive feedback by Dr. Thilo Womelsdorf. The manuscript in its current form was drafted by me and edited by Dr. Thilo Womelsdorf.

The research for **Chapter 4** was designed by myself, Dr. Anna Schubö and Dr. Thilo Womelsdorf. All data collection and analyses were performed by me. Dr. Marcus Watson helped to build the experimental setup. The manuscript was drafted by me and edited by all authors.