

NEURAL CORRELATES OF ADAPTIVE LEARNING IN THE CINGULATE CORTEX

Yin Li

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

Joshua I. Gold
Professor, Neuroscience

Graduate Group Chairperson

Joshua I. Gold
Professor, Neuroscience

Dissertation committee:

Yale Cohen, Associate Professor, Otorhinolaryngology
Nicole Rust, Assistant Professor, Psychology
Javier Medina, Associate Professor, Neuroscience
Michael Platt, Professor, Neuroscience

NEURAL CORRELATES OF ADAPTIVE LEARNING IN THE CINGULATE CORTEX

COPYRIGHT

2015

Yin Li

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit
<http://creativecommons.org/licenses/by-nc-sa/2.0/>

ACKNOWLEDGMENTS

I have many people to thank for the course of my scientific journey. First and foremost, I want to express my gratitude to my graduate adviser, Dr. Joshua Gold, whose steady hand has guided me through turbulent waters. He has done everything in his power to make my bumpy journey as smooth and as productive as possible. The Gold Lab has been a tremendously supportive environment. I want to acknowledge the help of lab members past and present, including Matt Nassar (who conceived of and continues to be an integral part of the adaptive learning project), Takahiro Doi (with whom I shared many recording sessions), David Barack (who prompted my exploration of the posterior cingulate), Sidd Joshi, Long Ding, Chris Glaze, Kamesh Krishnamurthy, Yunshu Fan, Rishi Kalwani, and Jeff Law. I would especially like to thank lab manager Jean Zweigle, who not only provided extraordinary care for our monkeys, but also provided a much needed perspective on life in the real world.

I want to thank the members of my thesis committee, Yale Cohen, Nicole Rust, Javier Medina and Michael Platt, who have been a source of encouragement and wisdom throughout my graduate career. The systems and computational neuroscience community at Penn has been both intellectually stimulating and generous with their time. The Neuroscience Graduate Program (embodied in Mikey Nusbaum, Jane Hoshi, and Christine Clay) and the MD-PhD Program (embodied in Maggie Krall and Skip Brass) provided necessary institutional support (and cheerleading).

A career in academia would not have been possible without the support of family and friends. My life has been a typical immigrant story with exceptional parents who sacrificed everything back home to ensure that I would have a successful future. Most of all, I thank my proof-reader, fellow scientist and fiancée—Anna Stern—who has been my anchor for the past few years as I roughed the seas of noisy, unpredictable experiments.

ABSTRACT

NEURAL CORRELATES OF ADAPTIVE LEARNING IN THE CINGULATE CORTEX

Yin Li

Joshua I. Gold

Learning from experience is critical for adaptive decision-making. When the world is unpredictable, the learning process itself must be adaptive. Outcomes that are attributed to a fundamental change in the environment should have a large impact on behavioral adjustment, whereas outcomes that are attributed to noisy, but expected, fluctuations in an otherwise stable environment should be largely ignored. There is growing interest in understanding the neural mechanisms of adaptive learning. However, studies to date have primarily used functional magnetic resonance imaging, which has limited temporal and spatial resolution. Relatively little is therefore known about how the brain implements adaptive learning at the level of individual neurons. Here, I advance the thesis that two reciprocally connected regions of the cingulate cortex—the anterior and the posterior cingulate cortex—contribute to adaptive learning by encoding environmental context and by encoding outcomes in a context-dependent manner. I support this thesis with recordings from individual neurons in the ACC and the PCC in monkeys performing an adaptive learning task.

TABLE OF CONTENTS

Acknowledgments	iii
Abstract	iv
List of Figures and Tables.....	vi
Chapter 1: Introduction	1
Chapter 2: Adaptive encoding of feedback by the cingulate cortex during adaptive learning	13
Chapter 3: Conclusions and Future Directions	43
References	60

LIST OF FIGURES AND TABLES

Figure 2-1: Change-point task and recording sites.....	32
Figure 2-2: Behavioral data from monkeys SP (a–c) and AT (d–f) and simulated data from a hybrid Bayesian model (g–i).....	33
Figure 2-3: Cingulate selectivity for feedback and context.	35
Figure 2-4: Context-dependent feedback encoding.	37
Supplementary Figure 2-S1: Adaptive integration of feedback history.	39
Supplementary Figure 2-S2: Model-based evidence for adaptive learning.	41
Supplementary Figure 2-S3: Heterogeneity and bias in neural selectivity.....	42
Figure 3-1: Actual and simulated behavior on the change-point task.....	56
Figure 3-2: Temporal dynamics of cingulate feedback responses.....	57
Figure 3-3: Two example ACC neurons with reward expectation signals.....	58
Figure 3-4: Cingulate neurons respond to surprising auditory stimuli.....	59

CHAPTER 1: INTRODUCTION

To make predictions about the future, we need to learn from our past. In an uncertain world, however, individual experiences are only partially predictive about future outcomes. A single cold winter does not dispel the notion of global warming. A single poor quarterly earnings report does not mean long-term decline in a company. And a single failed experiment does not discredit a graduate student's thesis. In each of these examples, expected noisy fluctuations in outcomes limit our interpretation of single experiences, and temper how much we should learn from them.

Expectations about statistical regularities in our environment are critical for appropriately interpreting experiences to adjust our beliefs and behaviors. This form of context-dependent process is known as adaptive learning. In this introduction, I review conceptual considerations regarding learning in probabilistic, dynamic environments, and why adaptive learning is required. I then turn to two areas in the cingulate cortex that may play a role in adaptive learning, with a focus on neuroanatomic and neurophysiologic findings that may have bearings on adaptive learning.

Learning in dynamic environments

To formalize the notion of "the extent of learning," let's first turn to reinforcement learning (RL) theory, which was independently developed in machine learning and animal psychology literature (Sutton and Barto, 1998).

According to the Rescorla-Wagner model of reinforcement learning, the expectation at time $t+1$ is given by updating the pre-existing expectation at time t (V_t) according to the difference between the actual outcome (R_t) and the pre-existing expectation (V_t):

$$V_{t+1} = V_t + \alpha_t(R_t - V_t) \quad (1)$$

where the difference $R_t - V_{t-1}$ is referred to as the prediction error (PE) and the constant of proportionality α_t is referred to as the learning rate. In classic learning experiments, R is usually a reward outcome and V is the expected reward (also known as the value) of a particular action or stimulus. However, the same framework can also be applied to more abstract problems, such as inferring the mean of a probability distribution (μ) that gives rise to observable outcomes (R) (Nassar et al., 2010).

In this framework, the learning rate serves to determine the relative weighting (or influence) of the new outcome versus past experiences (aggregated in terms of the pre-existing expectation). When the learning rate is 1, all of the weight is placed on the most recent outcome (i.e., $V_{t+1} = R_t$), whereas when the learning rate is 0, all of the weight is placed on the pre-existing expectation and no learning takes

place (i.e., $V_{t+1} = V_t$). For a constant learning rate between 0 and 1, the update equation performs a weighted sum of all previous outcomes, with the largest weight for the most recent outcome.

In order to ensure that V converges to the true reward rate rather than continuing to fluctuate in response to new outcomes even after the average reward rate has been learned, it is common in the machine learning literature to set the learning rate to the inverse of the number of outcomes that have been observed (Sutton & Barto, 1998). This procedure in essence takes an average of all the outcomes in an environment. As more and more data is accumulated, new information has decreasing influence on the expected reward rate of the environment.

In the real world, however, we often don't know when the environment has changed (and thus how many outcomes to average over). One solution for this problem is suggested by the Pearce-Hall model of learning, in which the learning rate itself depends on the absolute value of the prediction error (Pearce and Bouton, 2001). Such a model deals better with a dynamic environment because more unexpected or "surprising" outcomes lead to a higher learning rate, allowing the learner to also effectively discard past outcomes that are likely to be no longer relevant. In other words, the Pearce-Hall model uses the magnitude of the prediction error as a built in change-point detector.

A limitation of the Pearce-Hall model is that its conception of "surprise" fails to take into account the statistics of the environment. Consider, for example, a

monkey keeping track of how rewarding a fruit tree is. Finding no fruit when the typical yield is consistently 5 pieces of fruit is a lot more surprising than if the typical yield varies between 0 and 10 pieces of fruit, even if the average yield is 5 pieces of fruit both situations. How surprising an outcome is should thus depend on not just the mean of the reward distribution, but also its variance, what is referred to as "noise." It's the signal-to-noise ratio of a prediction error that matters for detecting environmental changes, not just the absolute value.

Another aspect of the environment that affects the interpretation of an outcome is the "volatility" of the environment. In a more volatile environment, where changes happen frequently, an unexpected outcome is more likely to result from a true change in the environment than if the environment were generally stable. For the foraging example, if a monkey knows that trees at a particular foraging site have a short life span, then finding 0 pieces of fruit should lead him to abandon that tree. On the other hand, if trees were long-lived, then he should be more persistent in returning to a tree that had yielded abundant fruit in the past. Noise and volatility thus play opposite roles in determining how much an animal should learn from the same prediction error: higher noise should suppress learning from an outcome, whereas higher volatility should enhance it.

These notions of noise and volatility can be formalized with a Bayesian framework that considers learning as a form of statistical inference. Estimating the average reward rate is an inference problem in which the underlying state of the

environment (μ or mean reward rate) is inferred from a noisy sequence of outcomes (R) that provide evidence about the current state (Yu and Dayan, 2004; Courville et al., 2006; Adams and MacKay, 2007; Behrens et al., 2007; Fearnhead and Liu, 2007; Nassar et al., 2010, 2012; Wilson et al., 2013). In this framework, the pre-existing expectation V_t from Eq. 1 can be thought of as a prior with distribution $P(\mu)$, the outcome R_t as the evidence with likelihood distribution $P(R|\mu)$, and the updated expectation V_{t+1} as the posterior $P(\mu | R)$. "Noise" is formalized as a type of uncertainty about the outcome, related to the width of the likelihood distribution. "Volatility" is formalized as a type of uncertainty about the prior, related to an overall "offset" that affects the likelihood that on average the environment could be in any other state. In reinforcement learning, the learning rate arbitrates between the relative influence of new outcomes and pre-existing expectations. In Bayesian learning, this arbitration occurs naturally as a function of the relative uncertainties for the prior versus the likelihood, by way of Bayes theorem:

$$P(\mu | R) \propto P(\mu)P(R|\mu). \quad (2)$$

When outcomes are noisy (wide likelihood), more weight is placed on the prior, and less learning occurs. When the average volatility is high, the prior is more uncertain, and more weight is placed on the outcomes, increasing learning. Moreover, recent changes in the environment also increase learning, since changes increase the uncertainty of a learner in its estimate of underlying environmental state (leading to wider priors). The Bayesian framework thus provides a principled means of

interpreting unexpected outcomes in a dynamic environment, and thus a normative account of how to adjust the learning rate.

Various names have been used to partition these types of uncertainties in the priors and likelihoods. In an influential review, Yu & Dayan (2005) coined the phrases "expected" and "unexpected uncertainties." Expected uncertainty arises from known unreliability in the relationship between outcome and the underlying state of the environment, and is thus related to noise. Unexpected uncertainty, by contrast, arises from outcomes that violate existing expectations about an environment, i.e., outcomes that are unexpected if the environment hadn't changed, and is thus related to both the volatility of an environment and the signal-to-noise ratio of an outcome. Yu & Dayan (2005) also noted, however, that expected uncertainty can be further divided into uncertainty arising from "ignorance" (which can be reduced by accumulating evidence) and "inherent stochasticity" in the task (related to noise), that cannot be reduced. The uncertainty from "ignorance" has been variously referred to as "estimation uncertainty" (Payzan-LeNestour et al., 2013) and "relative uncertainty" (Nassar et al., 2010), whereas the irreducible uncertainty has been referred to as "risk" (Bach and Dolan, 2012). Unexpected uncertainty has also been termed "change-point probability" (Nassar et al., 2010), "second-order uncertainty" (Bach and Dolan, 2012), and "surprise" (O'Reilly et al., 2013). A cautionary note is that the mappings between these terms are not always exact.

Regardless of what they are called, many recent studies have demonstrated that humans and animals are sensitive to these various forms of uncertainties and adapt learning accordingly (Gallistel et al., 2001; Daw et al., 2006; Behrens et al., 2007; Nassar et al., 2010, 2012; O'Reilly et al., 2013; McGuire et al., 2014; Diederer and Schultz, 2015). Behrens et al. (2006) found that human subjects adapted their learning rates according to the volatility of the environment. Nassar et al. (2010) found that human subjects adjusted their learning rates in a trial-by-trial fashion, as a function of both how noisy the environment is and how long they have been in a changed environment, which is related relative uncertainty. Diederer & Schultz (2015) likewise found that the sensitivity human subjects to reward prediction errors were scaled by reward variability. Even rodents are sensitive to environmental statistics: rats that are used to a volatile environment detected changes in reward contingencies more quickly than rats used to a stable environment (Gallistel et al., 2001).

Despite the abundance of behavioral evidence, the neural substrates underlying adaptive learning are poorly understood. The vast majority of the adaptive learning literature have used functional magnetic resonance imaging to probe the representation of these various forms of uncertainty and behavioral adaptation. However, blood oxygenation level dependent (BOLD) signals measured by functional imaging have limited spatial and temporal resolution (Logothetis, 2008). Moreover, BOLD signals do not map linearly onto neural responses (Logothetis et al., 2001). To understanding how the brain represents and supports

adaptive learning, it is thus important to directly record from neurons in humans or other animals as they perform adaptive learning. Below, I outline the rationale for doing such recordings in the cingulate cortex.

The cingulate cortex and adaptive learning

The cingulate cortex consists of a vast sheath of gray matter in the medial wall of the cerebral cortex that overlay the corpus callosum (Vogt and Pandya, 1987). The anterior cingulate cortex, which consists of Brodmann areas 24 and 32, has long been implicated in high-order cognitive function, including cognitive control and conflict-monitoring, pain and emotional processing, action selection and reward processing, although its exact function remains hotly contested (Ridderinkhof et al., 2004; Walton et al., 2007; Cole et al., 2009; Mansouri et al., 2009; Rushworth et al., 2011; Shenhav et al., 2013; Silvetti et al., 2014). The posterior cingulate cortex, which includes Brodmann areas 23 and 31, has traditionally been designated a node of the default-mode network, based originally on the striking activation of this area in functional imaging experiments when subjects are not performing tasks, although more recent work have focused on its role evaluating decision options and outcomes (McCoy et al., 2003; McCoy and Platt, 2005; Schacter et al., 2007; Hayden et al., 2008; Heilbronner and Platt, 2013; Leech and Sharp, 2013; Pearson and Platt, 2013). Several lines of evidence indicate that both the ACC and the PCC are suitable substrates for adaptive learning.

Neuroanatomical connectivity indicate that both ACC and PCC are nodes of highly connected hubs, allowing them to rapidly integrate sensory, motor and affective signals (Vogt et al., 1979, 2006; Vogt and Pandya, 1987; Heilbronner and Haber, 2014). In addition to their reciprocal connections with each other, both ACC and PCC are connected to the orbitofrontal cortex, which is implicated in value encoding, the dorsolateral prefrontal cortex, which is involved in working memory and executive function, and the hippocampal formation, which plays a key role in consolidation of long-term memory (Vogt and Pandya, 1987; Vogt and Gabriel, 1993; Heilbronner and Haber, 2014). The ACC is additionally connected to primary and association motor cortices, whereas the PCC receives strong projections from sensorimotor areas of the posterior parietal cortex (Vogt and Pandya, 1987; Heilbronner and Haber, 2014).

The innervation of the cingulate cortex by neuromodulatory systems bears special attention. Dopamine (DA), released by the midbrain neurons of the ventral tegmental area and substantia nigra pars compacta, are thought to encode reward prediction error signals that drive reinforcement learning (Schultz et al., 1997). Norepinephrine (NE), released by the pontine nucleus locus coeruleus, is involved in physiological arousal and has been proposed to signal unexpected uncertainty or change-point probability (Yu and Dayan, 2004; Aston-Jones and Cohen, 2005; Payzan-LeNestour et al., 2013). Acetylcholine (ACh), released by neurons in the basal forebrain as well as the brainstem, is involved in attention and learning and is thought to signal expected uncertainty or noise (Yu and Dayan, 2004; Lee and Dan,

2012). All three of these neuromodulatory systems send projections to the ACC (Vogt et al., 1979; Porrino and Goldman-Rakic, 1982; Heilbronner and Haber, 2014). Moreover, ACC is a primary cortical input to the locus coeruleus (Porrino & Goldman-Rakic, 1982). The connectivity of the PCC to neuromodulatory systems is less well documented, partly reflecting the relatively lesser attention that has been paid to this caudal cingulate area. However, metabotropic receptors for dopamine, acetylcholine and norepinephrine are all present in PCC (Palomero-Gallagher et al., 2009). Fibers that terminate in the PCC stain positive for enzymes required for catecholamines (DA and NE) and ACh biosynthesis, suggesting that PCC also receive inputs from these three neuromodulatory systems (Heilbronner et al., 2014). The ACC and PCC are thus well suited anatomically to integrate outcomes and contextual cues in mediating adaptive choices.

Single-unit recordings from monkeys are also consistent with a role for ACC and PCC in adaptive learning. Both regions encode actions and outcomes of actions, with subpopulations of neurons that encode positive and negative outcomes, prediction errors, and reward magnitude (McCoy et al., 2003; Amiez et al., 2005, 2006; Matsumoto et al., 2007; Hayden et al., 2008, 2009; Hayden and Platt, 2010; Kennerley et al., 2011). In ACC, a subpopulation of neurons encodes the surprisingness of outcomes, regardless of whether it is positive or negative, which may support a role in detecting environmental changes (Hayden et al., 2011). Both regions also encode the outcomes of previous trials, potentially allowing these areas to compare an outcome with the local statistics of rewards (Seo and Lee, 2007;

Hayden et al., 2008; Bernacchia et al., 2011; Kennerley et al., 2011). Finally, activity of these areas predict behavioral switching on tasks where switching represent a change in strategy or learning (Quilodran et al., 2007; Hayden et al., 2008).

Lesion experiments further highlight the importance of ACC and PCC in adaptive behavior. Notably, monkeys with ACC lesions were able to learn from errors on single trials, but could integrate feedback from across multiple trials to maintain optimal choices (Kennerley et al., 2006). Muscimol inactivation of ACC led to impaired performance in a probabilistic reversal task (Amiez et al., 2006). Muscimol inactivation of PCC impaired the ability to learn arbitrary visuo-saccadic associations when the expected rewards were low (but not when those rewards were high), suggesting a context-dependent role for the PCC in learning (Heilbronner and Platt, 2013).

Human neuroimaging studies have directly implicated ACC and PCC in adaptive learning. ACC activity has been found to track volatility, change-point probability, relative uncertainty, and model updating in subjects learning from probabilistic, dynamic environments (Behrens et al., 2007; Krugel et al., 2009; O'Reilly et al., 2013; McGuire et al., 2014). Moreover, ACC activity is stronger when feedback is more indicative of reversal in reward contingencies in a probabilistic reversal task, suggesting an adaptive coding function for the ACC (Jocham et al., 2009). PCC has similarly been implicated in the signaling of change-point probability and various forms of uncertainty (Payzan-LeNestour et al., 2013; McGuire et al., 2014).

In summary, neuroanatomical studies, recording and lesion experiments in monkeys, and neuroimaging studies in humans are consistent with roles for the ACC and the PCC in adaptive learning. Indeed, Pearson et al. (2011) has recently outlined a model whereby the ACC encodes task-relevant variables that are then integrated by the PCC to detect changes in the environment. Alternatively, the ACC and PCC might each encode contextual variables that are needed to drive adaptive learning. To test these hypotheses and better understand the contributions of the cingulate cortex in adaptive learning, I recorded from single neurons in the ACC and the PCC in two monkeys trained on a task developed by a previous graduate student in the lab, Matt Nassar, in which optimal performance required adaptive learning. The results of these experiments are presented in the main chapter of this dissertation.

CHAPTER 2: ADAPTIVE ENCODING OF FEEDBACK BY THE CINGULATE CORTEX DURING ADAPTIVE LEARNING

Yin Li, Matthew R. Nassar, Joseph W. Kable, Joshua I. Gold

Abstract

Effective error-driven learning in a dynamic environment must distinguish between different sources of errors, such as those arising from fundamental change-points that require behavioral adjustments versus those arising from noise that should be ignored (Yu and Dayan, 2004; Courville et al., 2006; Adams and MacKay, 2007; Fearnhead and Liu, 2007; Nassar et al., 2010; Wilson et al., 2013). Despite extensive evidence for representations of errors in the brain, little is known about if and how these representations are modulated by information about the source of the errors to guide effective learning behaviors. Here we show that individual neurons in both the anterior (ACC) and posterior (PCC) cingulate cortex, which can represent error-related feedback signals (McCoy et al., 2003; Behrens et al., 2007; Matsumoto et al., 2007; Seo and Lee, 2007; Heilbronner and Platt, 2013), can also: 1) encode contextual signals needed to distinguish different sources of those errors, and 2) use those signals to adaptively modulate error encoding, particularly in the PCC, in accordance with behavioral adjustments. We recorded from ACC and PCC neurons of monkeys that used normative, adaptive learning strategies to perform a

change-point task. Individual neurons in both regions were sensitive to error feedback and the relative balance of noise and change-points that contributed to the errors. Furthermore, in PCC but not in ACC, feedback encoding was adaptively modulated by context in a manner consistent with behavioral adjustments, such that errors related to change-points were enhanced relative to errors related to noise. These results suggest that cingulate-mediated, adaptive encoding of feedback may contribute to complex, adaptive learning behaviors.

Text

To examine the role of the cingulate cortex in adaptive learning, we recorded from ACC and PCC neurons in two monkeys performing a novel change-point task. The task required them to make a saccadic eye movement to predict the likely location of a single, stochastically determined "reward target" from among a circular array of ten visual targets, based on the outcomes of previous trials (Fig 1). The monkeys performed this task under two different, explicitly cued, blockwise conditions. In the zero-noise, high-change ("unstable") condition, reward was always paired with a single target that changed location relatively often (at a fixed "hazard rate" of 0.4–0.45 per trial). Thus, feedback was highly reliable. The best target to choose was always the previously rewarded target, even just after a change point. In contrast, in the high-noise, low-change ("noisy") condition, the location of the reward target fluctuated around a relatively stable best target that changed less

often (hazard rate=0.02). Thus, feedback was less reliable. Identifying the best target required integrating noisy feedback information across trials and often suppressing behavioral adjustments to small errors.

We analyzed the performance of the two monkeys from all sessions for which we have neural recordings ($n=93,278$ trials over 67 sessions for monkey SP; 71,422 trials over 65 sessions for monkey AT). Overall, the monkeys chose the reward target well above chance in both the unstable (49% for SP, 50% for AT) and noisy (42% for SP, 43% for AT) condition. The monkeys tended to choose the best target >50% of the time after just one trial after a change-point of the best target in the unstable condition and within 5–6 trials in the noisy condition (Fig. 2a,d). In the noisy condition, the monkeys reliably chose the best target more frequently than the reward rate for that target (60%, blue lines in Fig. 2a,d,g) within 10–15 trials after change-points and thus were not just choosing the most recent reward target or matching reward probabilities (Sugrue et al., 2004; Lau and Glimcher, 2005).

Both monkeys used adaptive, feedback-driven strategies to solve the task. They were more likely to switch choices after error versus correct feedback but did not treat all error feedback equally (Fig. 2b,e). Switching was more common when the spatial error magnitude (the distance from the reward target to the chosen target) was large. Moreover, for a given spatial error magnitude, the monkeys were less likely to switch in the noisy versus unstable condition. This difference was largest for small spatial errors, which were more predictive of change-points in the unstable versus noisy condition. Finally, for noisy trials, the monkeys were more

likely to switch during times of subjective uncertainty about the environment, as indexed by the variability of their choices (“choice entropy”) over the past ten trials (Fig. 2c,f).

These adaptive strategies, including sensitivity to noise condition and uncertainty, are consistent with human behavior on comparable change-point tasks and are predicted by normative (Bayesian) theory (Courville et al., 2006; Wilson et al., 2013; Gallistel et al., 2001; Daw et al., 2006; Behrens et al., 2007; Krugel et al., 2009; Nassar et al., 2010, 2012; O’Reilly et al., 2013; McGuire et al., 2014; Diederer and Schultz, 2015). The monkeys also exhibited non-normative tendencies to persist in choosing the same target. Accordingly, a hybrid model with fully adaptive, Bayesian learning plus persistent choices reproduced key features of the monkeys' behavior, including a dependence on error magnitude, noise, and choice entropy (Fig. 2g-i). This model provided a better fit to behavior than either a purely Bayesian model ($\Delta\text{BIC} = -113.48$ for SP, -121.77 for AT, paired t-test, $p < 0.001$ for both) or a standard, fixed-learning rate reinforcement learning model that also included the same tendency to persist ($\Delta\text{BIC} = -479.89$ for SP, -290.33 for AT, paired t-test, $p < 0.001$ for both).

To identify novel neural signals that underlie these adaptive strategies, we targeted the ACC and PCC. Both encode feedback at the single-neuron and population levels, and both are sensitive to relevant contextual information like change-point probability and subjective uncertainty about the current state of the environment at the population level (Vogt and Pandya, 1987; Ito et al., 2003;

Pearson et al., 2011; Heilbronner and Haber, 2014; Matsumoto et al., 2007; Quilodran et al., 2007; et al., 2008; Heilbronner and Platt, 2013; Behrens et al., 2007; Jocham et al., 2009; Payzan-LeNestour et al., 2013; Economides et al., 2014; McGuire et al., 2014). However, little is known about how the contextual information is represented and combined with feedback information by single neurons. We recorded the activity of 102 ACC units ($n=67$ from SP, 35 from AT) and 235 PCC units (75 from SP, 160 from AT) as the monkeys performed the change-point task (Fig. 1b).

Units in both areas encoded categorical feedback (correct or error, Fig. 3a,f), as reported previously (McCoy et al., 2003; Matsumoto et al., 2007; Quilodran et al., 2007; Hayden et al., 2008; Heilbronner and Platt, 2013). They also encoded the magnitude of the spatial error, a type of prediction error that strongly influenced switching behavior (Fig. 3b,g). Both kinds of feedback signals were present in similar fractions of units in both ACC and PCC (Fig. 3e,j), but they used different encoding schemes. ACC neurons were divided roughly equally in terms of higher firing rates for correct or error feedback or as a function of error magnitude (the median regression coefficient in the early feedback epoch for feedback $[-0.003]$ or error magnitude $[-0.056]$ did not differ significantly from zero, Wilcoxon signed-rank test $p=0.74$ and 0.51 , respectively). By contrast, PCC neurons were generally more active for error trials (median feedback coefficient $=-0.574$, $p<0.001$) and larger error magnitudes (median error magnitude coefficient $=0.132$, $p<0.001$).

ACC and PCC units were also selective for the contextual factors needed to evaluate the informativeness of the feedback for adaptive learning, including noise condition and choice entropy (Fig. 3c-e,h-j). Whereas feedback signals appeared transiently after the time of feedback onset, selectivity for noise condition and choice entropy tended to be stable throughout each trial, even for individual neurons (Spearman's correlation coefficient comparing regression coefficients in a 500-ms pre-fixation epoch and a 500-ms reward epoch for noise were 0.73 and 0.72 for ACC and PCC data, respectively, $p < 0.001$ in both cases, and for choice entropy were 0.59 and 0.60, respectively, $p < 0.001$ in both cases). Encoding of context at the single-unit level was heterogeneous, with no overall bias in encoding either noise condition or choice entropy (the median regression coefficient for either factor in the pre-fixation epoch did not differ significantly from zero for either ACC or PCC data, $p > 0.22$ in all cases). Thus, both ACC and PCC may play general roles in representing a diversity of feedback- and context-related information needed for adaptive learning.

Some individual ACC and PCC neurons also encoded combinations of factors that were key to adaptive learning, with a more systematic and behaviorally relevant representation in PCC. Specifically, the monkeys' behavior showed a (normative) decreased sensitivity to small errors in the noisy versus unstable condition (Fig. 2h). Neural encoding of feedback in certain individual cingulate neurons showed a similar effect. In the unstable condition, the example ACC unit in Fig. 4a fired more for error trials than correct trials just after feedback onset and

then more for correct trials a short time later. In the noisy condition, this same unit exhibited an attenuated response to error trials in the early epoch and an attenuated response to correct trials in the late epoch, consistent with the reduced informativeness of feedback in the noisy condition. The example PCC unit in Fig. 4b was more sensitive to error trials in the unstable versus noisy condition. Across the population, these feedback-noise interactions were represented differently in ACC and PCC. In the ACC, we found neurons with both enhancing and suppressing effects of noise on feedback encoding (fraction of units with significant interactions between feedback [correct vs error=1] and context [stable vs noisy] in the late feedback epoch=0.118 [95% CI=0.062, 0.20]), with no systematic relationship across the population (Fig. 4c,e,f). In the PCC, there also were examples of both enhancing and suppressing interactions (fraction of significant units=0.0851 [0.0528, 0.128]). However, across the population in PCC there was a consistent and systematic suppressive effect of noise condition on small errors (spatial magnitude=1 or 2) in the late feedback epoch (Fig. 4d,g,h), mirroring the behavioral effect of noise on switch (Fig. 2b,e).

These results suggest a selective role for PCC in integrating feedback and noise condition to encode errors adaptively. This adaptive feedback encoding mirrored the differential sensitivity of the monkeys to errors in the two noise conditions, with higher sensitivity to small errors in the unstable versus noisy condition. This sensitivity is consistent with the encoding of change-point probability, which also distinguishes errors that arise from change-points versus

noise, in BOLD PCC activation (Nassar et al., 2010, 2012; McGuire et al., 2014). Our results are thus in broad agreement with previous proposals that a key function of the PCC is to detect environmental changes (Pearson et al., 2011) and show for the first time how individual neurons encode key factors needed to carry out that function effectively in a dynamic environment.

In ACC, feedback-noise interactions were more heterogeneous and may reflect more diverse roles in error processing. Approximately half of these neurons encoded feedback more weakly in the noisy condition and more strongly in the unstable condition. This kind of encoding was similar to what we found for PCC and is reminiscent of surprise signals that have been reported for ACC neurons (Hayden et al., 2011). The remaining ACC neurons we found that were selective for these feedback-noise interactions encoded feedback more strongly in the noisy condition. This encoding may be related to conflict or cognitive control signals that have been ascribed to ACC, because feedback is more ambiguous and thus difficult to interpret in this condition (Sheth et al., 2012; Shenhav et al., 2013; Ebitz and Platt, 2015). The mixed encoding scheme of ACC neurons may also reflect a general role in representing feedback and contextual information that can be read out efficiently by downstream areas in a task-specific manner (Rigotti et al., 2013), possibly including the feedback-noise interactions represented in the PCC for this task.

More generally, further processing of these feedback- and context-related representations by downstream areas appears to be necessary to solve the task. We did not find consistent effects of high or low choice entropy (by median split) on

feedback encoding in either brain area, even though the monkeys interpreted errors differently according to choice entropy (H_0 : slopes of ROC [correct vs. error=1] for high- vs. low-choice entropy trials=1, $p>0.05$ by bootstrap for both ACC and PCC in both early and late feedback epochs). Moreover, the firing rate of neurons in the two areas were generally not predictive of whether the monkey would switch or not on the subsequent trial: when a binary variable indicating whether the monkey switched or stayed on the following trial was added as a co-regressor to a linear regression model that already includes feedback and context terms, the fractions of significantly predictive units were 0.088 [95% CI=0.041,0.161] in ACC and 0.111 [0.074,0.158] in PCC in the late feedback epoch.

Thus, the cingulate cortex encodes all of the feedback and contextual variables necessary for adaptive behavioral adjustments in our task. However, this region does not appear to be the final common pathway for the integration of feedback with context to drive behavior. Neither ACC nor PCC integrated feedback with subjective uncertainty (estimated by choice entropy), and neural activity in ACC and PCC was not robustly predictive of the behavioral adjustment on the subsequent trial. Other brain areas, such as the insula, orbitofrontal cortex, the posterior parietal cortex, and even further downstream oculomotor planning areas, may be more directly involved in the trial-by-trial adjustment of learning, as has been suggested by neuroimaging studies (O'Reilly et al., 2013; McGuire et al., 2014; Chau et al., 2015). Whether individual neurons in these brain areas encode or make use of adaptive feedback processing remains to be tested.

Methods

Subjects

Two adult male rhesus monkeys (*Macaca mulatta*) were used for this study. All training, surgery, and experimental procedures were performed in accordance with the National Institutes of Health's *Guide for the Care and Use of Laboratory Animals* and were approved by the University of Pennsylvania Institutional Animal Care and Use Committee.

Behavioral task

On each trial, the reward target was drawn probabilistically from a distribution centered on the best target. The best target was selected uniformly from the ten possible targets based on a change-point process governed by a flat hazard rate. In the low-noise, high-hazard condition, the best target was rewarded 100% of the time and the hazard rate was 0.4–0.45 per trial. In the high-noise, low-hazard condition (the "high noise" condition), the best target was rewarded 60% of the time, with adjacent pairs of targets rewarded at 15% and 5%, and the hazard rate was 0.02 per trial. Noise and hazard rate were counterbalanced to achieve similar overall performance for the two task conditions. The two environments were presented in blocks of 100–300 trials and cued explicitly by the color and size

of the fixation point (large red disk for "low noise" and small gray disk for "high noise").

Each trial began with the presentation of a central fixation point. The monkey had 500 ms to acquire fixation, after which the ten-target array (radius 10° visual angle) appeared. Monkeys needed to maintain fixation for a variable duration (~500–1000 ms) until the fixation point disappeared, signaling the monkey to indicate its prediction by making a saccadic eye movement to the predicted reward target. After the monkey had indicated its choice, visual feedback was given indicating the chosen target (center of the chosen target turned green) and the reward target (a large white target was overlaid on top of the reward target). Visual feedback was present for a variable duration (randomly chosen from 1200, 1350, 1500, 1650, and 1800 ms), during which time the monkey could freely view the screen. After the feedback was extinguished, the screen went blank, and, if the reward target had been chosen, the juice was delivered. The variable delay between feedback onset and juice delivery allowed us to differentiate between the effects of receiving feedback information and receive the juice reward itself.

Eye position was monitored using a video-based system (Eye-Link 2000; SR Research) sampled at 1000 Hz. Visual stimuli were generated using the Psychtoolbox (Brainard, 1997) and custom Matlab software and presented on an LCD monitor (Viewsonic) located 60 cm from the monkey's eyes.

Electrophysiology

Each monkey was implanted with a head holder and two recording cylinders that targeted the ACC and PCC on either the left side (monkey SP) or the right side (monkey AT). ACC cylinders were placed at Horsley-Clark coordinates 33 mm AP, 8 mm L for monkey SP and 43 mm AP, 8 mm L for monkey AT. PCC cylinders for both monkeys were placed at 0 mm AP, 5 mm L, and tilted at an angle of 8.5 deg along the ML plane to point downward toward the midline. For ACC recordings, we targeted areas 32/24 along the dorsal bank of the anterior cingulate sulcus (~4–6 mm below cortical surface). For PCC recordings, we targeted areas 31 and 23, in the posterior cingulate gyrus (~7–11 mm below cortical surface). Both brain regions were targeted using magnetic resonance imaging and custom software (Kalwani et al., 2009) as well as by listening for characteristic patterns of white and gray matter during recordings. Neural recordings were conducted using either single-contact glass-coated tungsten electrodes (Alpha-Omega) or multicontact linear electrode arrays (V-probe, Plexon) and a Multi-channel Acquisition System (Plexon). Spike waveforms were sorted offline.

Behavioral analyses

Choice entropy, a measure of behavioral variability that reflects decision uncertainty (Goñi et al., 2011; Takahashi et al., 2015), was computed using the monkey's choices in the 10 trials leading up to and including the current trial to generate an empiric probability distribution $P(k)$ for each target k . Choice entropy,

H_{choice} was then given, by analogy to entropy in information theory (Shannon, 2001), by:

$$H_{\text{choice}} = \frac{1}{\log_2(10)} \sum_k -P(k) \log_2(P(k)) \quad (1)$$

where the $1/\log_2(10)$ term scales H_{choice} between zero (minimally variable, if the same target was chosen in the last ten trials) and one (maximally variable, if choices were equally distributed among the ten targets).

Bayesian inference models

Optimal Bayesian inference in the change-point task involves computing the posterior distribution over the best target μ_t on trial t after observing the t -th reward target from the history of all reward targets $x_1, \dots, x_t$, $P(\mu_t | x_{1:t})$. We developed a recursive, online algorithm to compute this discrete posterior distribution (Adams and MacKay, 2007; Behrens et al., 2007; Fearnhead and Liu, 2007; Nassar et al., 2010; Wilson et al., 2013). Let H be the flat hazard rate determining whether the best target is the same as from the previous trial or repicked randomly from one of the ten possible targets. The prior probability over the best targets on trial t is then given by a weighted sum of the posterior on the best target from the previous trial and the flat uniform distribution, where H governs the weighting:

$$P(\mu_t | x_{1:t-1}) = P(\mu_{t-1} | x_{1:t-1})(1-H) + \frac{1}{10}H. \quad (2)$$

The posterior probability $P(\mu_t | x_{1:t})$ after observing the t -th outcome x_t is then given by Bayes' Rule:

$$P(\mu_t | x_{1:t}) = \frac{P(x_t | \mu_t)P(\mu_t | x_{1:t-1})}{P(x_t | x_{1:t-1})}, \quad (3)$$

where $P(x_t | \mu_t)$ is the known likelihood or noise distribution and $P(\mu_t | x_{1:t-1})$ acts as a normalization term. Substituting in Eq. 3 gives a recursive algorithm for computing the posterior on the best target, given knowledge of the hazard H and the noise distribution $P(x_t | \mu_t)$:

$$P(\mu_t | x_{1:t}) = \frac{P(x_t | \mu_t) \left(P(\mu_{t-1} | x_{1:t-1}) (1-H) + \frac{1}{10} H \right)}{P(x_t | x_{1:t-1})}, \quad (4)$$

where $P(\mu_{t-1} | x_{1:t-1})$ equals $1/10$ for $t = 1$.

For the fitted Bayes model, we assumed that the probability of picking target k on the $t + 1$ trial was given by the softmax transformation of the posterior probability on the t -th trial:

$$P_{\text{Bayes}}(k, t+1) = \exp(\beta \times P(\mu_t = k | x_{1:t})) / \sum_k \exp(\beta \times P(\mu_t = k | x_{1:t})) \quad (5)$$

where β was an inverse-temperature term, with $\beta=0$ resulting in all targets being chosen with equal probability and higher β increasing the probability of picking the most probable target. We modeled the (discrete) noise distribution after the von-Mises distribution for circular variables (Fisher, 1995):

$$P(x_t | \mu_t) = \exp\left(\kappa \cos\left(\frac{\pi}{5}(x_t - \mu_t)\right)\right) / \sum_{\mu_t} \exp\left(\kappa \cos\left(\frac{\pi}{5}(x_t - \mu_t)\right)\right) \quad (6)$$

where the parameter κ is inversely related to the width of the distribution. The fitted Bayes model had five free parameters (two noise-dependent hazard rates H , two noise-dependent inverse-noise parameters κ , and the inverse temperature term β).

The hybrid Bayes model used exactly the same Bayesian inference algorithm but with three additional terms that described the non-normative tendency of the monkeys to persist in picking the same target (stay) after correct feedback (s_{pos}), error feedback in the low-noise condition ($s_{\text{neg,low}}$), and error feedback in the high-noise condition ($s_{\text{neg,high}}$). The probability of picking target k on the t -th trial was then given by:

$$P(k, t) = P(\mu_{t-1}=k | x_{1:t-1}) \times (1-s) + s \quad \text{if } k \text{ was chosen on the } t-1 \text{ trial} \quad (7)$$

$$P(k, t) = P(\mu_{t-1}=k | x_{1:t-1}) \times (1-s) \quad \text{if } k \text{ was not chosen on the } t-1 \text{ trial}$$

where $P(\mu_{t-1}=k | x_{1:t-1})$ is the posterior on best targets as computed in Eq. 5 in the standard Bayes model, and s is the noise- and feedback-dependent stay term that enhances the probability of the monkey to choose the same target as the previous trial. The hybrid Bayes model thus had seven free parameters (two noise-dependent hazard rates H , two noise-dependent inverse-noise parameters κ , and three noise- and feedback-dependent stay terms s). Note that the hybrid model did not have an additional inverse temperature term.

Reinforcement-learning models

We applied a reinforcement learning (RL) model to the monkeys' trial-by-trial choice behavior (Sutton and Barto, 1998). The value $V_k(t)$ of a visual target k on trial t was updated according to the following delta rule:

$$V_k(t+1) = V_k(t) + \alpha(R_k(t) - V_k(t)) \quad (9)$$

where R_k is 1 if target k was the reward target and 0 otherwise, and α is a learning rate that differed for the two models as follows. The learning rate α differed depending on whether or not k was the chosen target: the chosen target was associated with an "experienced learning rate" and all non-chosen targets were associated with a "fictive learning rate" (Hayden et al., 2009). The probability of choosing target k , $P(k)$, for our "hybrid" RL model was given by the softmax transformation of V_k (as in Eq. 5) followed by choice heuristics (as in Eqs. 7-8). The hybrid RL model had six free parameters (two learning rates, β , and three stay terms s).

Model fitting and comparisons

All models were fit using Matlab's "fmincon" function to find values of free parameters that maximized the likelihood of the monkeys' actual sequences of choices for each session. To compare goodness-of-fit for the models while accounting for differences in the number of free parameters, we used Bayesian information criterion (BIC ;Burnham and Anderson, 2004):

$$\text{BIC} = -\log L + m \log n, \quad (10)$$

where L is the likelihood of the monkey's choices given model parameters, m is the number of model parameters (seven for hybrid Bayes, five for pure Bayes, six for hybrid RL), and n is the number of trials. Lower BIC is associated with better fits.

Neural analyses

Neurons were included in the database for further analyses only if we recorded ≥ 350 total trials, of which ≥ 125 were low-noise trials and ≥ 150 were high-noise trials. This selection criterion yielded a database consisting of 102 units from ACC ($n=67$ from SP, 35 from AT) of which 77 were considered high-isolation quality single-units ($n=49$ from SP, 28 from AT), and 235 units from PCC ($n=75$ from SP, 160 from AT) of which 188 were considered high-isolation quality single-units ($n=58$ from SP, 130 from AT). On average, 710 trials (308 low noise, 402 high noise) were recorded for ACC units and 832 trials (329 low noise, 503 high noise) were recorded for PCC units. Isolation quality did not affect our main conclusions (i.e., whether an effect is significant or not), so here we present analyses that include both low- and high-isolation quality units.

Multiple linear regression

For our main regression analyses, we fit trial-by-trial firing rate, FR, of individual neurons to the following equation:

$$FR = \beta_0 + \beta_1 * fdbk + \beta_2 * errMag + \beta_3 * noise + \beta_4 * H_{choice} \quad (11)$$

where $\beta_0 - \beta_4$ are regression coefficients fit by Matlab's "glmfit," "fdbk" denotes feedback on the current trial (0 = error, 1 = correct); "errMag" is the spatial error

magnitude (mean-centered for error trials to decorrelate this regressor with respect to fdbk); "noise" denotes the noise condition (0 = low noise, 1 = high noise); and H_{choice} is the choice entropy (mean-centered for each noise condition to decorrelate this regressor with respect to noise). To determine whether neural on the current trial predicts switching behavior on the subsequent trial, we added "switch" (0 = switch, 1 = stay) as a coregressor. Significant selectivity was based on a nested F -test comparing the full model with reduced models that lacked the individual terms (equivalent to Student's t -test, $p < 0.05$).

For the sliding-window analyses (Fig. 3e,j), we performed the above regression for firing rates computed in 250-ms time bins advanced in 50 ms steps from 2500 ms before feedback onset to 2500 ms after feedback onset. For epoched analysis, we defined a pre-fixation epoch (500 ms before fixation point onset), an early feedback epoch (250-750 ms after feedback onset), a late feedback epoch (750-1250 ms after feedback onset), and a reward epoch (500 ms after feedback offset, when reward would be delivered).

ROC-slope analyses

We estimated the strength of feedback encoding by computing receiver-operator characteristic (ROC) indices based on the distributions of firing rates for correct trials versus error trials (Fig. 4). To determine whether context (e.g., noise condition or choice entropy) affected the strength of feedback encoding across the

population of neurons, we related ROC indices for one context (e.g., high noise) to another (e.g., low noise) using Deming regression, which is appropriate when there is measurement error in both variables, and reported the slopes of the resulting regression lines (Cornbleet & Gochman, 1979). To examine the effect of noise on feedback encoding, we computed slopes separately for each error magnitude. To examine the effect of choice entropy on feedback encoding, we computed slopes only using trials from the noisy condition where the error magnitude was 0 or 1.

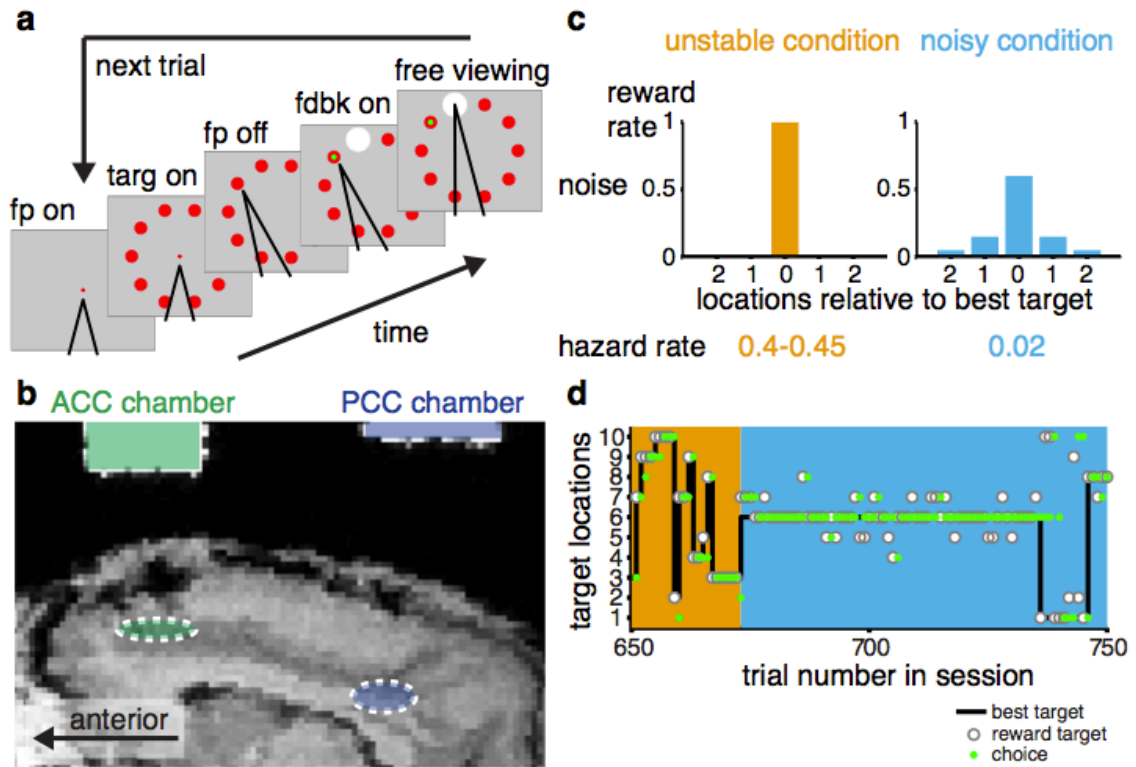


Figure 2-1: Change-point task and recording sites.

a On each trial, the monkey chose one of ten targets with a saccade. Visual feedback was then given for 1200–1800 ms, indicating the chosen target (small green dot) and the reward target (large white disk). Juice was then delivered if the monkey chose the reward target. **b** Structural MRI for monkey SP (~7 mm L), showing recording chambers (shaded rectangles) and approximate recording locations (dashed; determined as in {Kalwani et al 2009}). **c** The two task conditions, determined by: *i*) the hazard rate (H) governing the rate at which the best target was repicked at random from among the ten targets and *ii*) the noise in the distribution of reward probability as a function of distance from the best target. **d** An example sequence of trials for monkey AT. Background color indicates the task condition, as in **c**.

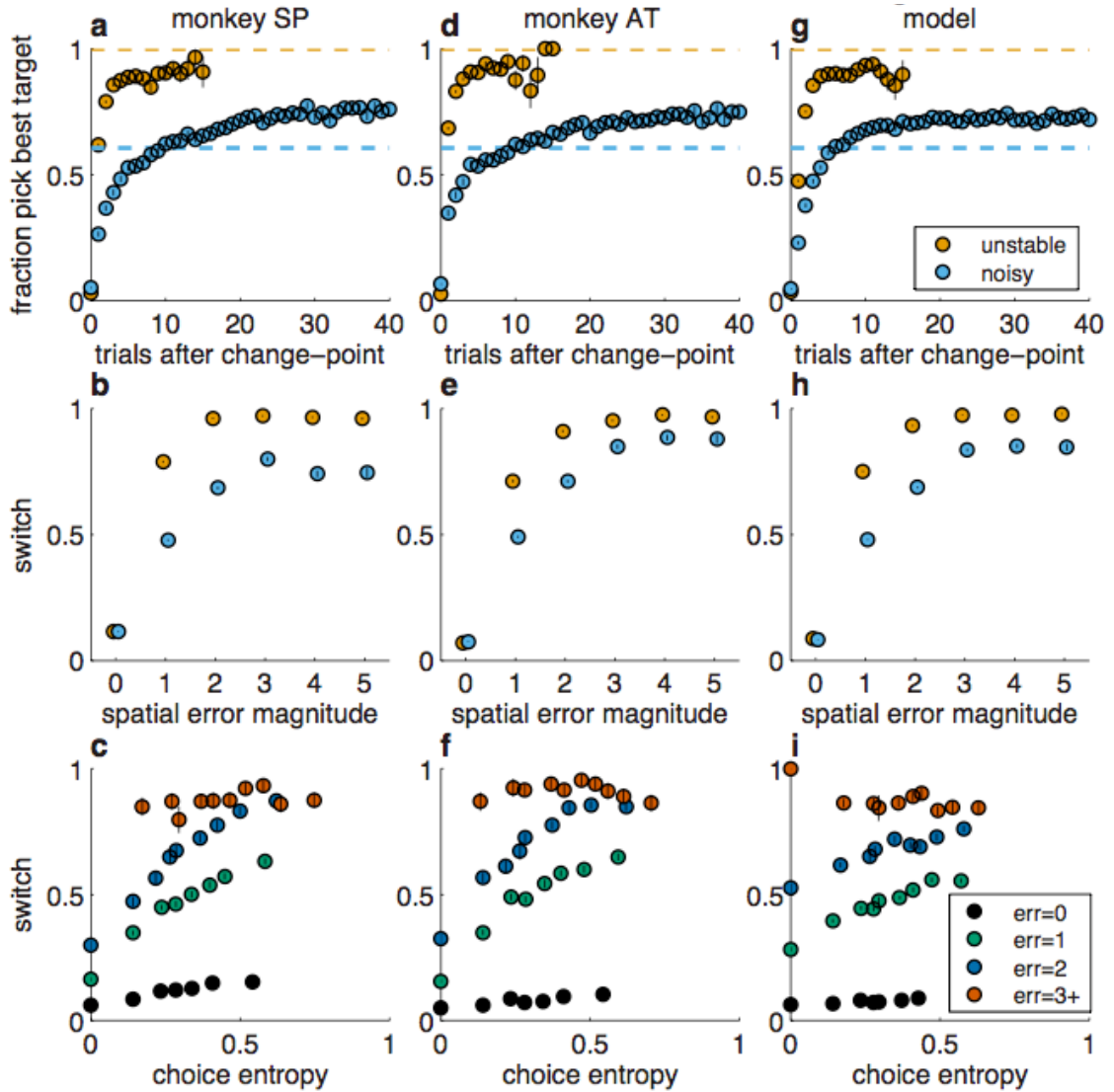


Figure 2-2: Behavioral data from monkeys SP (a-c) and AT (d-f) and simulated data from a hybrid Bayesian model (g-i).

a,d,g Fraction of trials picking the best target as a function of trials after change-point, plotted separately for unstable (yellow) and noisy (blue) conditions. Dashed lines indicate the fraction of trials in which the best target was rewarded for each task conditions. **b,e,h** Probability of switching to a different target on the following trial, per condition, as a function of the number of target locations between the chosen target and the reward target. Switch probability depended on spatial error magnitude (Spearman's $\rho=0.62$ for SP and 0.67 for AT, $p<0.001$ for both) and task condition (partial Spearman's ρ accounting for spatial error magnitude= -0.15 for SP, -0.10 for AT, $p<10^{-16}$ for both), particularly for small errors (median difference for errors of 1 and 2= -0.31 versus 3+ errors= -0.17 for SP, sign-rank test $p<10^{-10}$; -0.24

versus -0.09 for AT, $p < 10^{-10}$). Colors as in **a**. **c,f,i** Probability of switching in the noisy condition as a function of choice entropy computed from the previous ten trials (not including the first ten trials after change-points), plotted for different error magnitudes. Switch rate depended on choice entropy (partial ρ , accounting for error magnitude=0.21 for SP, 0.22 for AT, $p < 10^{-10}$ for both). In all panels, points/error bars indicate mean/sem across trials.

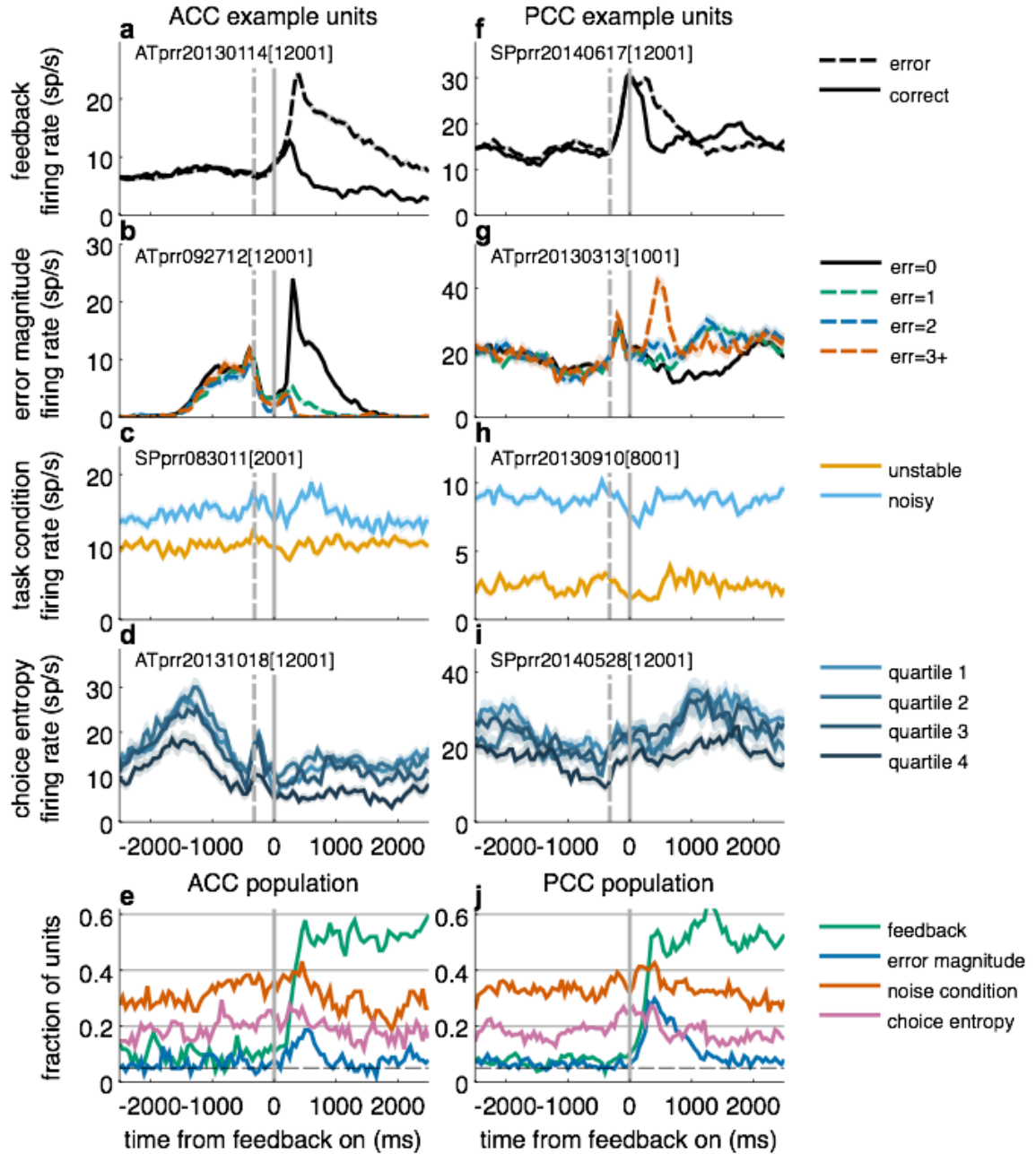


Figure 2-3: Cingulate selectivity for feedback and context.

a-d,f-i Example units from ACC (left column) and PCC (right column) illustrating neural selectivity for: **a,f** feedback (correct versus error trials); **b,g** spatial error magnitude; **c,h** task condition; and **d,i** choice entropy, grouped by quartiles (e.g., quartile 1 = 0-25 percentile) and only including trials from the noisy condition. Thick lines/ribbons indicate mean/sem firing rates in 100-ms sliding windows. Dashed vertical gray line indicates median time of saccade onset. **e,j** Population

selectivity for ACC ($n=102$ units) and PCC ($n=235$ units) for the indicated task variables, based on nested F -test with multiple linear regression computed on spike counts in 250 ms bins stepped in 50 ms increments ($p<0.05$). All panels show data aligned to feedback onset.

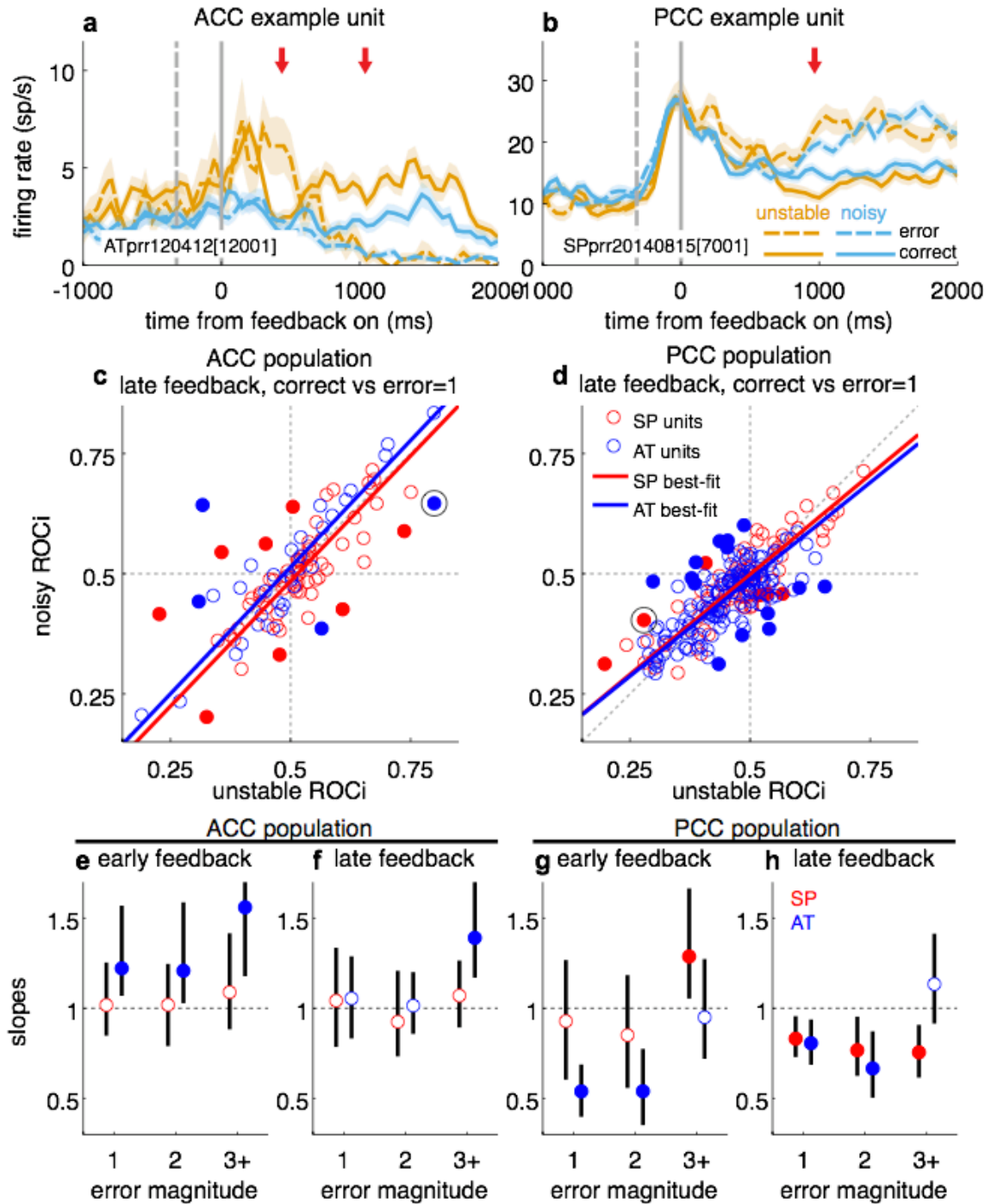
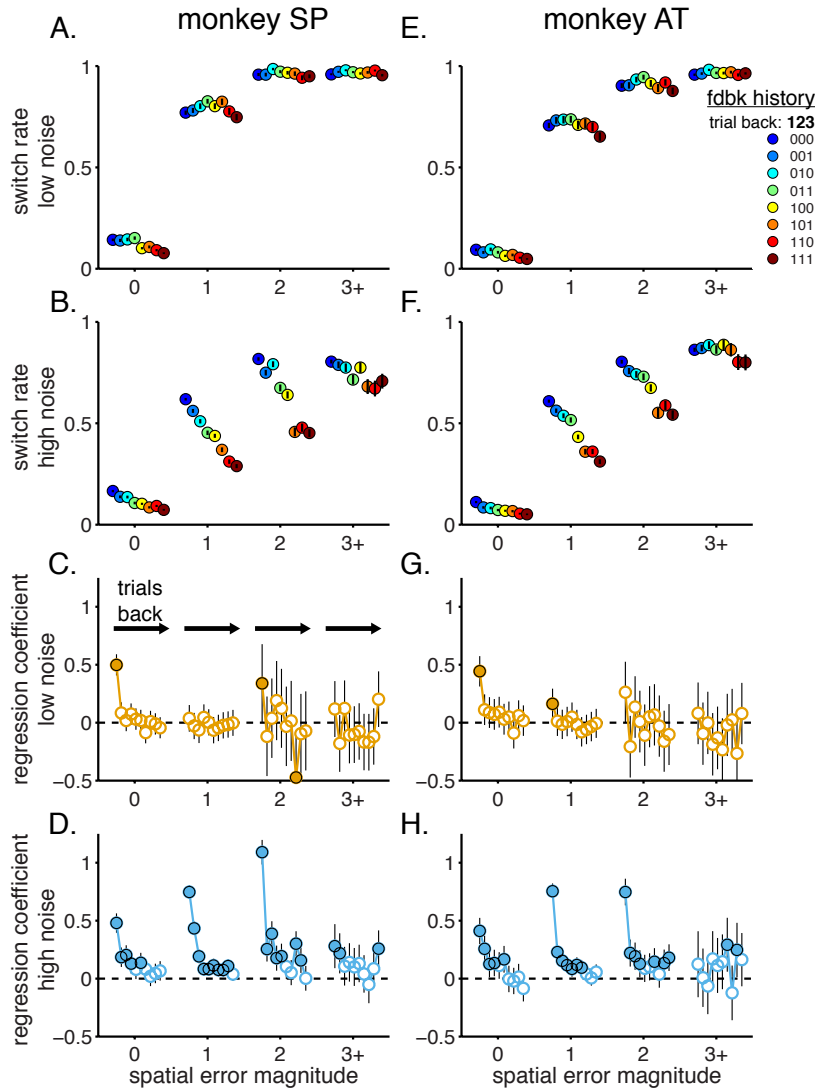


Figure 2-4: Context-dependent feedback encoding.

a Example ACC unit, exhibiting negative feedback encoding in the unstable (no-noise) but not the noisy condition in the early feedback epoch (first red arrow from left) and positive feedback encoding that was attenuated in the noisy versus

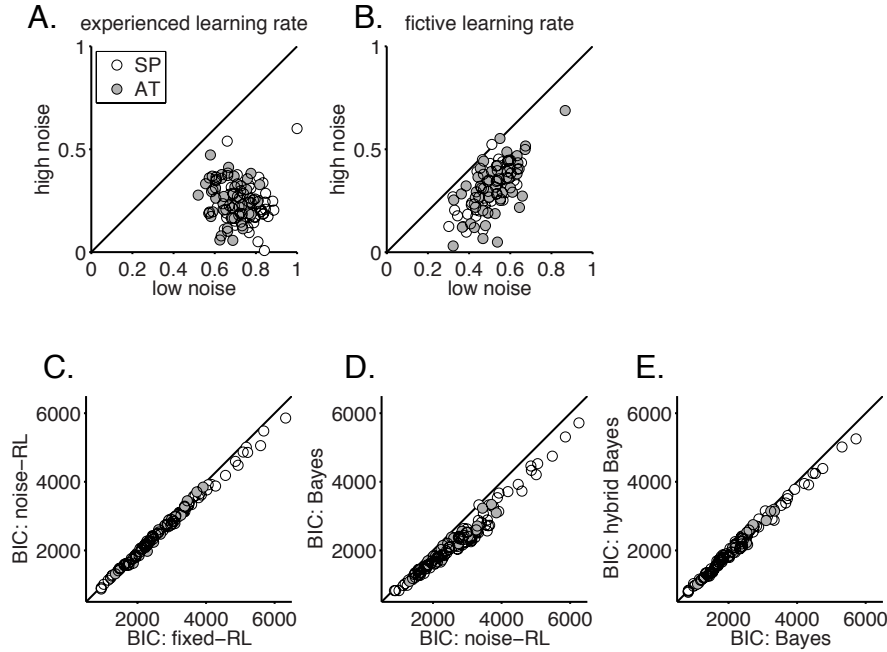
unstable condition in the late feedback epoch (second red arrow). **b** Example PCC unit with negative feedback encoding that was attenuated in the noisy versus unstable conditions. For **a,b**, only trials with spatial error magnitude of 0 (correct) or 1 (error) were included. **c,d** Population activity in the late feedback epoch for ACC (**c**) and PCC (**d**). ROC indices (ROCi) indicate whether neural activity was larger for correct (ROCi>0.5) or error=1 (ROCi<0.5) trials, separately per unit (points) for unstable and noisy trials. Filled symbols indicate units for which ROCi significantly differs for unstable versus noisy trials ($p<0.05$ by permutation test). Black circles indicate the example units in **a** and **b**. Best-fit lines for each monkey, as indicated, have slopes <1 ($p<0.05$ by bootstrap) for PCC units, indicating that noise had a suppressive effect on the strength of feedback encoding for these units. **e,h** Summary of context-dependent feedback encoding for ACC (**e,f**) and PCC (**g,h**) populations. Slopes of noisy versus unstable ROCi and corresponding 95% bootstrap confidence intervals are plotted for each monkey and error magnitude, as indicated, in the early- (250–750 ms, **e,g**) and late- (750–1250 ms **f,h**) feedback epochs. Filled symbols indicate conditions in which slopes deviated significantly from unity ($p<0.05$ by bootstrap).



Supplementary Figure 2-S1: Adaptive integration of feedback history.

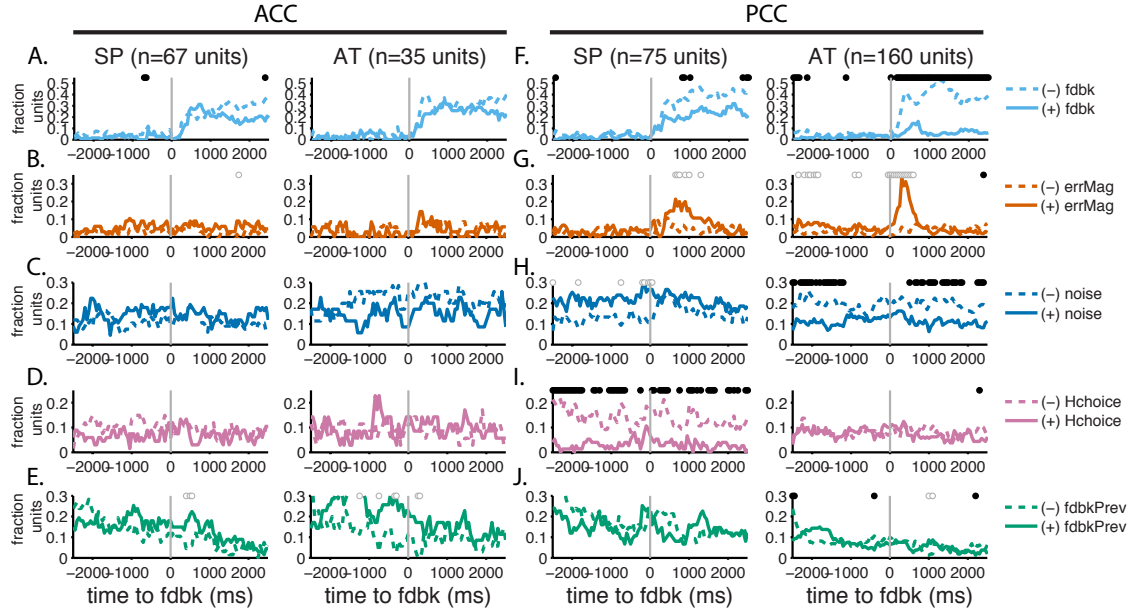
(A-D) Data for monkey SP. (E-H) Data for monkey AT. (A,E) Probability of the monkey switching for unstable (non-noisy) trials. For each spatial error magnitude on the current trial (spatial errors of 3 and greater are grouped together as 3+), behavior is separately plotted for different combinations of feedback history for up to 3 trials back. In legend, 0 = error and 1 = correct such that 000 (blue) indicates three consecutive error trials preceding the current trial whereas 001 (light blue) indicates two consecutive error trials preceded by a correct trial. Error bars = SEM. (B,F) Probability of the monkey switching for noisy trials, color coding as in (A,E). (C,G) Logistic regression coefficients for the influence of previous error trials on switching after the current trial in unstable trials. For each error magnitude, the influence of error feedback from up to 10 trials back is plotted (going from left to

right) from more recent to more distant trials. Positive coefficients imply that an error from that many trials back promotes switching (above what is expected for that particular error magnitude and noise condition). Error bars = 95% confidence interval. Significantly non-zero coefficients are filled in. (D,H) Same conventions as (C,G), but for noisy trials. These results indicate that the monkeys adaptively integrated feedback across multiple trials when feedback was ambiguous (e.g., small errors in noisy trials), but not when it was strongly predictive of change-points (e.g., large errors in noisy trials or any error in unstable trials).



Supplementary Figure 2-S2: Model-based evidence for adaptive learning.

(A, B) Fitted learning rates from the noise-adaptive reinforcement learning model. Experienced learning rates are used to update the value of the chosen target whereas fictive learning rates are used to update the values of the nonchosen targets. High-noise learning rates were lower than low-noise learning rates in every session (for experienced learning rates: paired Student's t -test, $p < 10^{-45}$ for SP, $p < 10^{-37}$ for AT; for fictive learning rates, $p < 10^{-33}$ for SP, $p < 10^{-25}$ for AT). White and gray disks symbols indicate fits from behavioral sessions with monkey SP and AT, respectively. (C) BIC values for the noise-adaptive reinforcement learning model (noise-RL) were lower than BIC values for a fixed-learning rate reinforcement learning model (fixed-RL) in 129 of 132 behavioral sessions, implying a better fit for the noise-adaptive model. (D) BIC values for the fitted Bayes model were lower than BIC values for the noise-adaptive RL model in every behavioral session. (E) BIC values for a hybrid Bayes model with non-normative tendency to stay after wins and losses were better than BIC values for the purely Bayes model in 118 of 132 behavioral sessions, indicating that the monkeys used a combination of normative and non-normative strategies.



Supplementary Figure 2-S3: Heterogeneity and bias in neural selectivity.

For each brain area and monkey, fraction of units with significant selectivity are plotted separately for significant positive encoding (solid line) and negative encoding (dashed line) for each listed factor, using same regression model as in Figure 6. Small circles at the top each panel indicate time bins where fraction of significant units are biased for either positive (open gray) or negative (filled black) encoding (binomial test, H_0 : proportion = 0.5, $p < 0.05$). For feedback encoding (A,F) and previous feedback encoding (E,J), positive encoding means firing more for correct trials. For spatial error magnitude (B,G), positive encoding means firing more for larger error magnitudes. For noise condition (C,H), positive encoding means firing more for high noise trials. For choice entropy (H,I), positive encoding means firing more for higher choice entropy.

CHAPTER 3: CONCLUSIONS AND FUTURE DIRECTIONS

In the previous chapter, I showed that monkeys trained to make predictions in a dynamic environment exhibit key features of adaptive learning, treating feedback differently depending on environmental and internal contextual cues. Neurons from both ACC and PCC encoded trial-by-trial feedback, contextual factors needed to interpret the feedback, and context-modulated feedback signals. Here, I discuss some of the implications of the study, relationship to other work, open questions and ongoing/future work to address them.

Neural mechanisms for adaptive learning

We identified two potential mechanisms whereby ACC and PCC can contribute to adaptive learning. First, individual ACC and PCC neurons encode contextual cues that are important for adaptive interpretation of feedback. Such contextual signals could be read out by other areas, such as the orbitofrontal cortex or the dorsolateral prefrontal cortex, to influence the computation of option values and policy selection (Lee and Seo, 2007; Rangel et al., 2008; Rushworth et al., 2011; Cai and Padoa-Schioppa, 2012). Alternatively, these contextual cues could be jointly computed in a distributed frontoparietal network or reflect coordinated input from neuromodulatory systems, such as cholinergic and noradrenergic systems, which are thought to broadcast expected and unexpected uncertainty, respectively (Yu

and Dayan, 2004; Aston-Jones and Cohen, 2005; Payzan-LeNestour et al., 2013). It has been proposed that ACC activity reflects volatility (Behrens et al., 2007; Rushworth and Behrens, 2008). However, because of our task design, in which noise and hazard rate were counterbalanced to minimize differences in difficulty between the two noise conditions, we were unable to dissociate coding for noise versus volatility.

Second, ACC and PCC neurons exhibit context-dependent modulation of feedback encoding. In the case of PCC, the modulation of feedback by noise was exactly of the form that would lead to adaptive response to errors in noisy conditions. Such context-dependent feedback encoding could be a general mechanism for adaptive learning in many brain areas. For example, in one neuroimaging study, human subjects were exposed to monetary gains and losses in a "win" context, where possible outcomes ranged from large monetary gain to no gain, and a "lose" context, where possible outcomes ranged from no loss to large monetary loss (Nieuwenhuis et al., 2005). Reward-sensitive brain areas, including the striatum, the prefrontal cortex, the cingulate, and the inferior parietal lobule, were as responsive to the largest gain in the "win" context as they were to the lack of loss in the "lose" context, suggesting that feedback signals were range-adapted. Even dopaminergic neurons exhibit context-dependent prediction errors, a topic that I will return to later.

Whereas PCC neurons were relatively homogeneous in their responses, with the majority of neurons firing more for error trials, larger spatial error magnitudes, consistent with a role in change detection, ACC neurons were more heterogeneous in the way they responded to change-points. The function of this heterogeneity is unclear, although it has been hypothesized that mixed encoding of task variables can be efficiently readout in downstream areas (Rigotti et al., 2013). One possibility is that ACC neurons also signaled change-points, but through shifts in ensemble activity. Such a mechanism has been reported for the medial prefrontal cortex (a homolog of the primate ACC) in rodents performing a reversal task (Karlsson et al., 2012). They found that the time at which rodents detected a change in reward contingency was preceded by a change in the ensemble activity of medial prefrontal neurons (which they referred to as a "network reset") while maintaining the overall firing rate of the neural population. Such a network reset would not be detectable by our single-cell recording technique. Toward the end of my recordings, I started to use multi-contact recording electrodes. It would be interesting to see if shifts in ensemble activity could be detected in these simultaneously recorded small populations of ACC and PCC neurons. More generally, for large and heterogeneous brain areas such as the cingulate cortex and complicated tasks such as the change-point task where no two trials are exactly the same, large-scale, simultaneous recording of dozens of neurons will likely yield novel insights about neural computations of learning by leveraging the power of single-trial decoding strategies (Cohen and Maunsell, 2010, 2011; Kiani et al., 2014; Kaufman et al., 2015).

Normative and non-normative behaviors

Behavior of the monkeys in the change-point task matched the qualitative predictions of adaptive learning models: decreased sensitivity to errors in the high-noise condition and increased sensitivity in regimes of high uncertainty. Moreover, a Bayesian model fit behavior better than a reinforcement learning model with noise-dependent learning rates, suggesting that the monkeys' behaviors were better captured by a model that involves trial-by-trial adaptive learning. However, the monkeys also behaved suboptimally. In the low-noise condition, when the monkeys should have always switched in response to any error, they nevertheless stayed on a substantial fraction of trials ($\sim 30\%$ for both monkeys). In the high-noise condition, by contrast, the monkeys switched more frequently than optimal for small errors and stayed more frequently than optimal for large errors. Fitting the noise and hazard rate parameters of the Bayesian model to the monkeys' trial-by-trial behavior confirms that these suboptimal behaviors can be explained at least in part by a mis-estimation of the statistics of the environment. The fit hazard rates were higher and the fit noise widths were wider than those used by the generative distribution. Similar mis-estimations of environmental statistics have also been found in human subjects (Nassar et al., 2010). Given that the monkeys were trained for many months on this task, the mis-estimation of environmental statistics is somewhat puzzling and may reflect underlying neurobiological constraints.

Even accounting for any mis-estimation of environmental statistics, the monkeys choice behavior were still influenced by non-normative factors. A model that included both Bayesian inference and non-normative tendencies for the monkeys to stay after wins and losses proved a better fit. However, we do not know the precise factors that led to this non-normative behavior. One possibility is that the problem is not with the inference (i.e., learning) but with the action selection. Most implementations of reinforcement learning models consists of both a "Critic" (which is involved in learning the values of various actions and stimuli) and an "Actor," which is involved in choosing the right action given the value functions (Sutton and Barto, 1998) . In our models, we have modeled the actor as simply a soft-max transformation of the posterior on the best target. But it is possible that the monkeys used a more complex or history-dependent policy. It's also possible that the monkeys were doing a form of off-line policy updating, so that even when a change-point had been detected, it is not immediately reflected in the action.

An alternative possibility is that the apparently non-normative behavior reflects model-free reinforcement learning acting in concert with Bayesian model-based learning mechanisms. This hypothesis is motivated in part by the observation that the monkeys' behavior (Fig. 1A) exhibits features of both a simulated RL model (Fig. 1B) and a Bayesian model (Fig. 1C). The hybrid model proposed in Chapter 2, which contains additional terms that reflect fixed tendencies to repeat choices, is shown in Fig. 3.1D for comparison. Most notably, the RL model "naturally" exhibits a tendency to repeat choices, whereas the Bayes model does not. Such an idea is

consistent with previous work indicating that both model-based Bayesian and model-free RL systems can be operating simultaneously in human subjects and that the relative uncertainty of the two systems can arbitrate between whether one or the other is in control at any given time (Daw et al., 2005; Gläscher et al., 2010; Lee et al., 2014) . Ongoing collaboration with a postdoctoral fellow now at Columbia, David Barack, is aimed at disentangling the relative influences of various learning and choice models to more accurately capture the monkeys' strategies in this task.

Generality of the findings

Several factors potentially limit the generality of our findings to adaptive learning in "the real world." First, steady-state performance on this complex task required months of training for both monkeys. We do not know if the signals that we observed in the cingulate cortex had been present from the beginning, or only emerged after months of training as has been observed in certain types of perceptual learning (Law and Gold, 2008) . Humans and animals often learn in rapidly changing environments without the need to train extensively (although they still need to learn or be instructed of the environmental statistics). It is not clear whether or how learning about the task may have influenced the representation of learning-related signals in ACC and PCC.

Second, the functional homology between human ACC and monkey ACC has been the subject of some dispute. For example, conflict-related activity has long

been found in human ACC, including single-unit recordings from human subjects, but such activity has been much harder to find in the monkey ACC (Cole et al., 2009; Mansouri et al., 2009; Silvetti et al., 2014) . Evidence for conflict-related activity in monkey ACC has recently emerged, although only of a specific type, suggesting that the conflict-related activity in ACC may be more restricted than in humans (Ebitz and Platt, 2015) .

Third, ACC and PCC are large brain areas that are likely to exhibit significant heterogeneity across its extent. Single-unit electrophysiology experiments likely sample from slightly different portions of the cingulate in different studies and different labs. Even between our two monkeys, we observed some differences in neural activity for the same brain area, which may be related to differences in the sampling of neurons or individual differences between the two monkeys. In this respect, the ability to sample the entire brain (as in neuroimaging) can be quite useful.

To address these concerns as well as to replicate our and previous findings, we are collaborating with a graduate student in Joe Kable's lab, Arthur Lee, to implement the same monkey task in human subjects. Interestingly, preliminary results suggest that at least some human subjects also exhibit non-normative tendencies to repeat choices in the high-noise condition. Ongoing work is aimed at collecting a robust database of human subject behavior and brain scans to see if our single-unit results replicate.

Temporal dynamics of cingulate responses

We observed a number of interesting temporal dynamics in cingulate responses that deserve further analysis and investigation. The overall latency to feedback response ~ 500 ms was similar in both ACC and PCC. A dissociation emerged, however, when positive and negative feedback were separately considered. In ACC, positive encoding emerged first, at ~ 500 ms after feedback onset, and this was followed by the emergence of negative feedback encoding at ~ 750 ms (Fig. 2A). By contrast, PCC exhibited the opposite pattern, with negative encoding (~ 500 ms) preceding positive encoding at ~ 750 ms. The relative timing remained the same when we considered the time to peak selectivity of positive or negative encoding in the two areas (Fig. 2B). Simultaneous recordings from the two areas, perhaps including local field potential activity, would be needed to confirm this relationship. If confirmed, these preliminary results suggest that ACC and PCC might receive different inputs. One possibility is that ACC receives an early positive prediction error signal from ascending DA neurons, whereas PCC receives an early negative prediction error signal from the habenula (Schultz et al., 1997; Bromberg-Martin et al., 2010). Given the reciprocal connectivity between ACC and PCC, the later feedback responses could reflect reciprocal communication between the two brain areas. Although highly speculative, this line of reasoning suggests that analyzing cingulate responses separately for positive and negative encoding might be useful.

Within ACC, we also identified a number of neurons with striking temporal dynamics related to the time of juice delivery. Recall that the task had imposed a variable delay between feedback onset and juice delivery, such that the monkey could anticipate approximately but not exactly when juice would be delivered. The example neuron in Fig. 3A fired more for correct trials than error trials within ~250 ms after feedback onset. It then fired a phasic burst after feedback offset (and reward onset) whose amplitude depended on the feedback duration. When feedback duration was shorter than average (blue, cyan traces), the neuron fired more strongly, whereas when the duration was longer than average (yellow, orange traces), the neuron fired more weakly. This response is reminiscent of the temporal difference learning signal that has been observed in midbrain DA neurons (Schultz et al., 1997) .

The second example neuron, in Fig. 3B, exhibited ramping activity starting around 1000 ms after feedback onset and persisted until ~200 ms after feedback offset and reward onset, after which the neural response was sharply suppressed. This response can be thought of as an reward expectancy response and bears similarities to phasic ramping responses that have been previously reported for ACC, although in that report, reward expectancy grew across multiple trials (Shidara and Richmond, 2002) . Over our entire database of recordings, we identified only a handful of neurons that exhibited these patterns, all of which were in the ACC. Since only a few neurons were found, we did not analyze them any further. However, it is tempting to speculate that the temporally discounted reward

signal in Fig. 3A could be shaped by an antagonistic relationship with neurons exhibiting the anticipatory ramping response in Fig. 3B.

An important future direction is therefore characterizing and modeling the temporal dynamics of the cingulate responses to understand when and why context-dependent signals emerge on a cell-by-cell basis. A number of computational models have been proposed to describe temporal dynamics in the ACC (Alexander and Brown, 2011; Silvetti et al., 2011) . Do these models adequately describe the diversity of the neurons in ACC and PCC?

Task relevant and irrelevant surprise signals: relationship to the locus coeruleus

In the change-point task, we found that both ACC and PCC contained subsets of cells that responded to the spatial error magnitude of the feedback, a form of task-relevant surprise signal. In a related project (led by postdoc Sidd Joshi) characterizing the relationship between pupil activity and various cortical and subcortical areas (including ACC, PCC, and LC), we found that neurons in the two areas also responded to unexpected auditory stimuli during fixation (Fig. 4). Notably, cingulate responses included a phasic burst that peaked ~250 ms after "beep" onset and a more sustained response that lasted up to 1000 ms; the tone itself was only 500 ms long. These findings suggest that cingulate neurons respond to (1) both surprise signals that required learning (task-relevant) and surprise

signals that were uninformative (task-irrelevant surprise signals), and (2) both visual and auditory signals that are surprising (i.e., multimodal).

These task-irrelevant surprise signals have also been found in the locus coeruleus, although with notably shorter duration of responses lasting < 500 ms (Joshi et al., submitted). In human subjects performing an adaptive learning task, playing a surprising (but task-irrelevant) auditory stimulus was found to evoke both pupil responses and increased learning of task-relevant stimuli (Nassar et al., 2012). This was interpreted as having an effect on the locus coeruleus, which in turn broadcast a surprise signal to the rest of the brain in the form of norepinephrine release (Yu and Dayan, 2004; Aston-Jones and Cohen, 2005). These sets of findings suggest that one source of the surprise signals we found in ACC and PCC may be LC activity. Alternatively, those signals may be computed in the cingulate cortex and transmitted to the LC. Sidd Joshi, the same fearless postdoc who has already ventured into the LC in two previous studies, is planning to imminently return to the LC and the ACC for dual recordings to address some of these questions.

Adaptive coding of feedback: relationship to dopamine

Our results are consistent with the hypothesis that a subset of neurons in ACC and PCC encode not just the prediction error, but one scaled by context-dependent learning rates (i.e., the update signal). An alternative possibility, however, is that the prediction error is already context-dependent and

appropriately scaled within the mesolimbic dopamine signal when it arrives as an input. Support for this view comes from single-unit electrophysiology experiments in monkeys indicating surprisingly sophisticated encoding of prediction errors by the DA system (Nakahara et al., 2003; Tobler et al., 2005; Bromberg-Martin et al., 2010a) . In the first study to show context-dependent encoding of prediction errors by dopamine neurons, Nakahara et al. (2003) found that when reward on one trial predicted lack of reward on the next (because of probabilistic structure of the task), DA neurons correctly decreased their expectation of reward on the trial following a rewarded trial. By contrast, in a task where there was no such dependence, DA neurons increased their expectation of rewards after each rewarded trial, consistent with classic reinforcement learning theory. In another study, monkeys were cued not just about the average magnitude and probability of juice rewards (i.e., expected value), but also the range of rewards (Tobler et al., 2005). They found that the prediction errors in midbrain DA neurons scaled with the range of expected reward magnitudes. More recently, Bromberg-Martin et al. (2010) found that both midbrain DA neurons and the lateral habenula (a major input) were sensitive not only to experienced rewards but also inferred value of a non-rewarded target based on task-specific statistical relationship between the rewarded and non-rewarded target.

These findings defy a simple parsing of the brain's learning systems into a mesolimbic model-free system and a cortical model-based system (Daw et al., 2005). Indeed, recent neuroimaging work suggest that both the DA system and

more "cognitive" networks, including the cingulate cortex, have access to context-dependent or model-based reward signals (Nieuwenhuis et al., 2005; Daw et al., 2011). In the case of our task, it would be extremely interesting to record from midbrain DA neurons and the lateral habenula to examine the relationship between subcortical and cortical encoding of prediction errors to see when and how context-dependence emerges.

Concluding remarks

In an uncertain world where individual experiences are only partially predictive of the future, adaptive learning allows subjects to rapidly adjust their expectations according to the statistical context through which experiences are interpreted. The results of my experiment show that the anterior and posterior cingulate cortex encode correlates of the contextual cues needed to adjust this interpretational process as well as context-dependent feedback signals that in some cases reflect adaptive interpretation of feedback by monkeys. The causal role of ACC and PCC and their relative contributions to adaptive learning remain to be determined, but these results add to the growing evidence that the cingulate cortex may play important roles in the adaptive adjustment of behavior through learning.

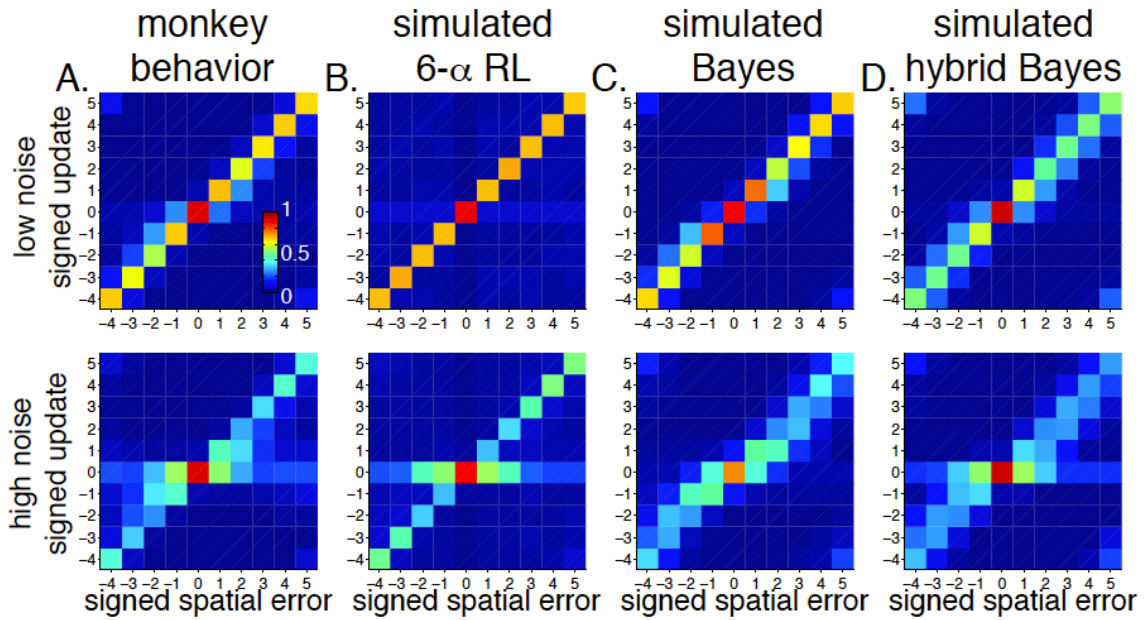


Figure 3-1: Actual and simulated behavior on the change-point task.

(A) Heatmaps showing the frequency of the monkey's update (spatial difference between choice on next trial and choice on current trial) for each signed spatial error (spatial difference between reward target and choice on current trial) for low-noise trials (top) and high-noise trials (bottom). Data is collapsed across all trials for both monkeys. For (B-D), simulated model behavior using the median fit parameters and the same sequence of outcomes experienced by the monkeys. (B) Simulated behavior for RL model with six learning rates, three for each noise condition: a "positive" learning rate associated with the chosen target on correct trials, a "negative" learning rate associated with the chosen target on error trials, and a "fictive" learning rate associated with the non-chosen targets. This model also had an inverse temperature term (total 7 free parameters). (C) Simulated behavior for "pure" Bayes model with noise and hazard rates fitted to the monkey's behavior. (D) Simulated behavior for "hybrid" Bayes model with fitted noise/hazard rates as well as 1 win-stay and 2 noise-dependent lose-stay terms. The monkeys' behavior (A) shares features of both RL (B) and Bayes models (C), especially in terms of a tendency to "stay" on high-noise trials (horizontal "band" at update=0).

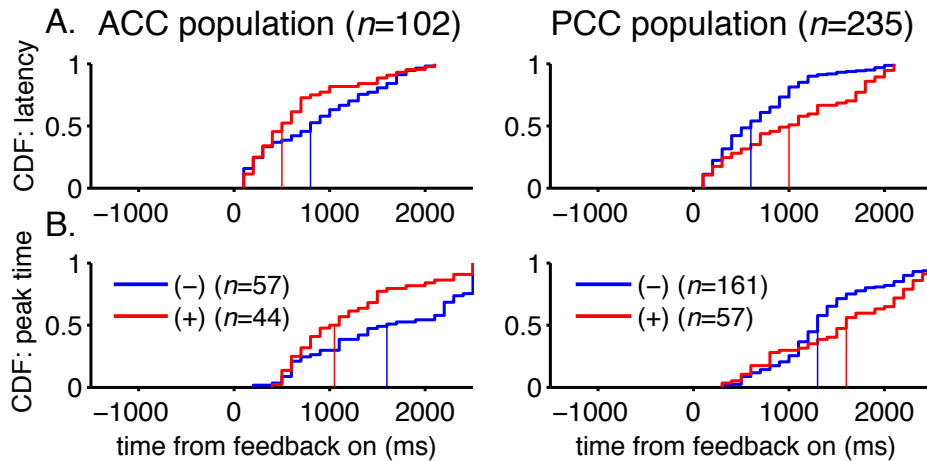


Figure 3-2: Temporal dynamics of cingulate feedback responses.

Cumulative density functions for (A) the latency and (B) peak time of feedback selectivity for ACC (left) and PCC (right) neurons. Selectivity based on linear regression analysis computed on firing rates in 250-ms sliding windows in 50-ms steps. A neuron must have at least 5 contiguous bins after feedback onset with $p < 0.05$ (Student's t -test) to be considered feedback-selective. (A) Latency computed based on first time bin of at least 5 contiguous, selective time bins. Positive (red) and negative (blue) encodings are plotted separately. A single neuron can contribute to both curves. (B) Time to peak selectivity based on time when absolute value of the t -statistic is maximum, subject to the 5 contiguous-bin criterion. For ACC, positive encoding precedes negative encoding (two-sample Kolmogorov–Smirnov test, $p < 0.05$, for both A and B). For PCC, positive encoding follows negative encoding (two-sample Kolmogorov–Smirnov test, $p < 0.05$, for both A and B).

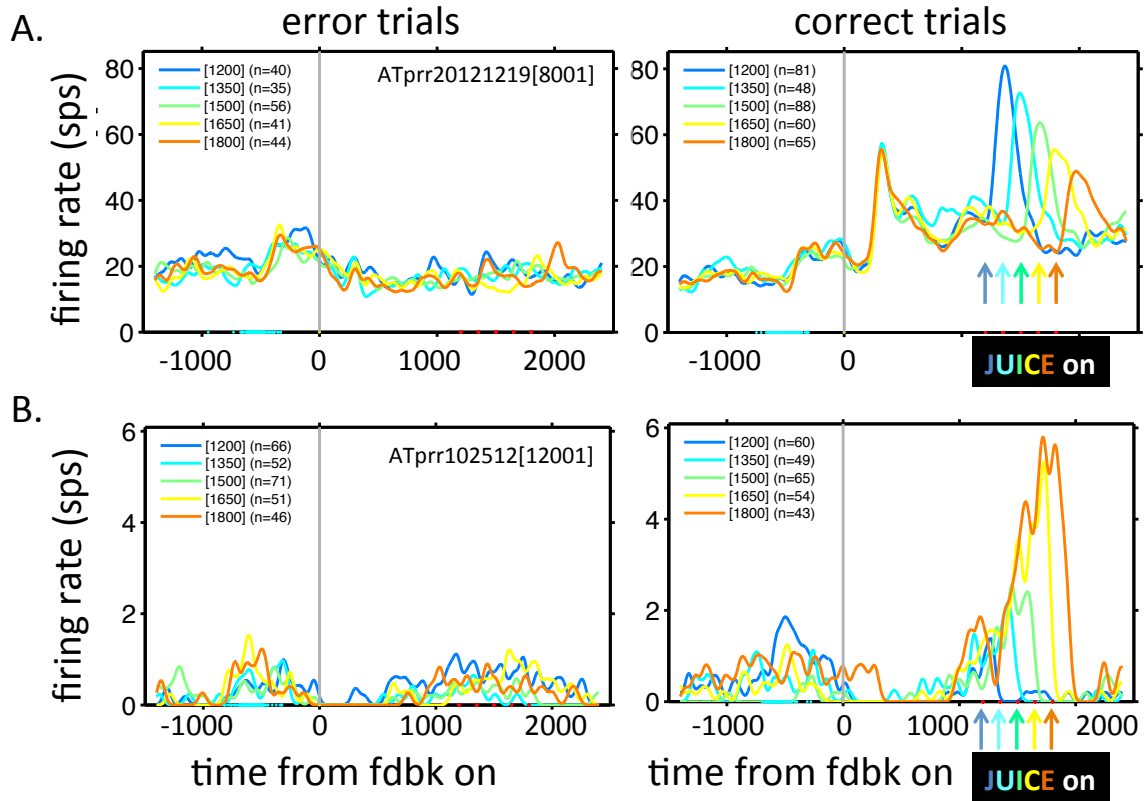


Figure 3-3: Two example ACC neurons with reward expectation signals.

For both neurons, firing rate is plotted separately for error (left) and correct (right) trials, for each feedback duration (from 1200 ms, blue to 1800 ms, orange). Arrows indicate the time of juice delivery. (A) Positive feedback encoding neuron that responds more strongly to feedback offset when it is earlier than average, and less strongly to feedback offset when it is later than average, reminiscent of temporal difference signal. (B) Positive feedback encoding neuron with ramping response that begins ~1000 ms after feedback onset and is terminated within ~200 ms after feedback offset, reminiscent of a reward anticipation signal.

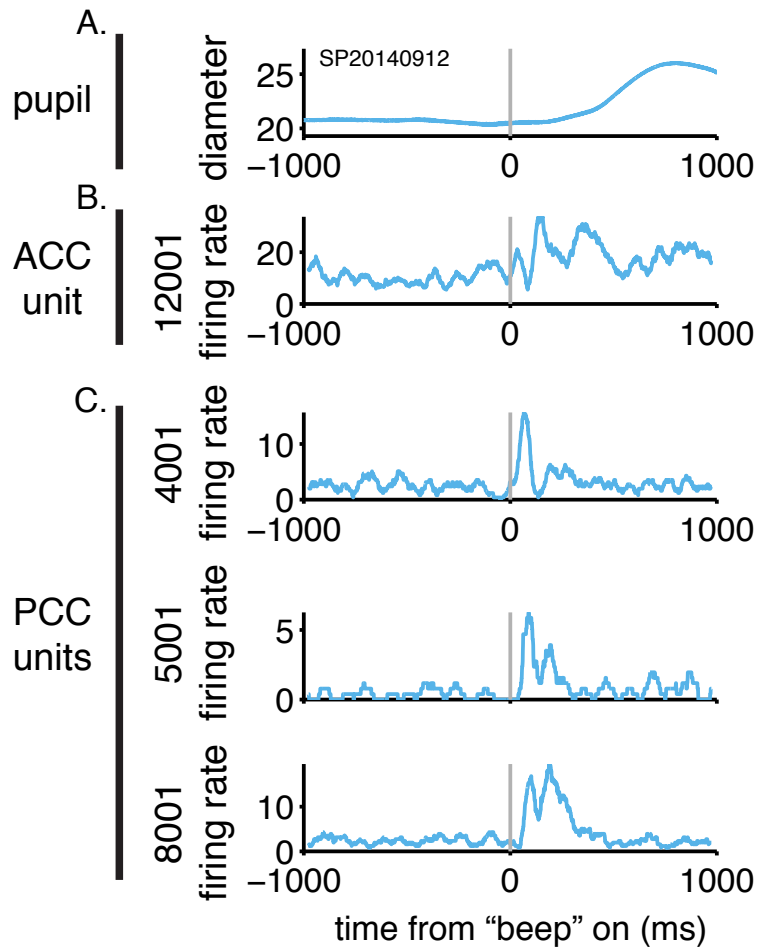


Figure 3-4: Cingulate neurons respond to surprising auditory stimuli.

Simultaneously recorded (A) pupil diameter (arbitrary units), (B) ACC response (sps), and (C) PCC responses (sps) to an unexpected auditory tone ("beep") that was presented on 25% of fixation trials in a single session for monkey SP.

REFERENCES

- Adams RP, MacKay DJC (2007) Bayesian online changepoint detection. arXiv preprint.
- Alexander W, Brown J (2011) Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience* 14:1338–1344.
- Amiez C, Joseph J, Procyk E (2005) Anterior cingulate error-related activity is modulated by predicted reward. *European Journal of Neuroscience* 21:3447–3452.
- Amiez, Joseph JP, Procyk (2006) Reward Encoding in the Monkey Anterior Cingulate Cortex. *Cerebral Cortex* 16:1040–1055.
- Aston-Jones G, Cohen J (2005) AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annu Rev Neurosci* 28:403–450.
- Bach D, Dolan R (2012) Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat Rev Neurosci* 13:572–586.
- Behrens T, Woolrich M, Walton M, Rushworth M (2007) Learning the value of information in an uncertain world. *Nature Neuroscience* 10:1214–1221.
- Bernacchia A, Seo H, Lee D, Wang X-J (2011) A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience* 14:366–372.
- Bornstein A, Daw N (2013) Cortical and Hippocampal Correlates of Deliberation during Model-Based Decisions for Rewards in Humans. *PLoS Computational*

Biology 9.

Brainard DH (1997) The psychophysics toolbox. Spatial vision Available at:

<http://bbs.bioguider.com/images/upfile/2006-4/200641014348.pdf>.

Bromberg-Martin E, Matsumoto M, Hong S, Hikosaka O (2010a) A Pallidus-

Habenula-Dopamine Pathway Signals Inferred Stimulus Values. *Journal of Neurophysiology* 104:1068–1076.

Bromberg-Martin E, Matsumoto M, Nakahara H, Hikosaka O (2010b) Multiple

Timescales of Memory in Lateral Habenula and Dopamine Neurons. *Neuron* 67.

Burnham KP, Anderson DR (2004) Multimodel inference understanding AIC and

BIC in model selection. *Sociological methods & research* Available at:

<http://smr.sagepub.com/content/33/2/261.short>.

Cai X, Padoa-Schioppa C (2012) Neuronal Encoding of Subjective Value in Dorsal and

Ventral Anterior Cingulate Cortex. *The Journal of Neuroscience* 32:3791–3808.

Cohen M, Maunsell J (2010) A Neuronal Population Measure of Attention Predicts

Behavioral Performance on Individual Trials. *J Neurosci* 30:15241–15253.

Cohen M, Maunsell J (2011) Using Neuronal Populations to Study the Mechanisms

Underlying Spatial and Feature Attention. *Neuron* 70.

Cole M, Yeung N, Freiwald W, Botvinick M (2009) Cingulate cortex: diverging data

from humans and monkeys. *Trends in neurosciences* 32:566–574.

Courville A, Daw N, Touretzky D (2006) Bayesian theories of conditioning in a

changing world. *Trends in Cognitive Sciences* 10:294300.

Daw N, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal

- and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* 8:1704–1711.
- Daw N, O'Doherty J, Dayan P, Seymour B, Dolan R (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
- Dean H, Crowley J, Platt M (2004) Visual and Saccade-Related Activity in Macaque Posterior Cingulate Cortex. *Journal of Neurophysiology* 92:3056–3068.
- Diederen K, Schultz W (2015) Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology*:jn.00483.2015.
- Doll B, Duncan K, Simon D, Shohamy D, Daw N (2015) Model-based choices involve prospective neural activity. *Nature Neuroscience* 18:767–772.
- Donahue C, Seo H, Lee D (2013) Cortical Signals for Rewarded Actions and Strategic Exploration. *Neuron* 80.
- Ebitz B, Platt M (2015) Neuronal Activity in Primate Dorsal Anterior Cingulate Cortex Signals Task Conflict and Predicts Adjustments in Pupil-Linked Arousal. *Neuron* 85:628640.
- Economides, Guitart-Masip, Kurth-Nelson, Dolan (2014) Anterior Cingulate Cortex Instigates Adaptive Switches in Choice by Integrating Immediate and Delayed Components of Value in Ventromedial Prefrontal Cortex. *Journal of Neuroscience* 34:33403349.
- Fearnhead P, Liu Z (2007) On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*

69:589–605.

Fisher NI (1995) Statistical analysis of circular data. Available at:

<http://books.google.com/books?hl=en&lr=&id=wGPj3EoFdJwC&oi=fnd&pg=PA1&dq=fisher+1993+statistical+analysis+of+circular+data&ots=PjZwvyBHSb&sig=gbGZDOvPooRSnK5XwMeZGDhFSNc>.

Gallistel, Mark T, King A, Latham (2001) The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes* 27:354.

Gläscher J, Daw N, Dayan P, O'Doherty J (2010) States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron* 66.

Hampton A, Bossaerts P, O'Doherty J (2006) The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. *The Journal of Neuroscience* 26:8360–8367.

Hayden B, Heilbronner S, Pearson J, Platt M (2011) Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *The Journal of Neuroscience* 31:4178–4187.

Hayden B, Nair A, McCoy A, Platt M (2008) Posterior Cingulate Cortex Mediates Outcome-Contingent Allocation of Behavior. *Neuron* 60.

Hayden B, Pearson J, Platt M (2009) Fictive Reward Signals in the Anterior Cingulate Cortex. *Science* 324:948–950.

Hayden B, Platt M (2010) Neurons in Anterior Cingulate Cortex Multiplex

- Information about Reward and Action. *The Journal of Neuroscience* 30:3339–3346.
- Heilbronner S, Haber S (2014) Frontal Cortical and Subcortical Projections Provide a Basis for Segmenting the Cingulum Bundle: Implications for Neuroimaging and Psychiatric Disorders. *The Journal of Neuroscience* 34:10041–10054.
- Heilbronner S, Platt M (2013) Causal Evidence of Performance Monitoring by Neurons in Posterior Cingulate Cortex during Learning. *Neuron* 80.
- Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* Available at: <http://www.sciencemag.org/content/302/5642/120.short>.
- Jocham G, Neumann J, Klein T, Danielmeier C, Ullsperger M (2009) Adaptive Coding of Action Values in the Human Rostral Cingulate Zone. *The Journal of Neuroscience* 29:7489–7496.
- Kalwani R, Bloy L, Elliott M, Gold J (2009) A method for localizing microelectrode trajectories in the macaque brain using MRI. *Journal of Neuroscience Methods* 176:104111.
- Karlsson M, Tervo D, Karpova A (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338:135–139.
- Kaufman MT, Churchland MM, Ryu SI, Shenoy KV (2015) Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. *eLife* Available at: <http://elifesciences.org/content/4/e04677.abstract>.
- Kennerley S, Behrens T, Wallis J (2011) Double dissociation of value computations

- in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience* 14:1581–1589.
- Kennerley S, Walton M, Behrens T, Buckley M, Rushworth M (2006) Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience* 9:940–947.
- Kiani R, Cueva CJ, Reppas JB, Newsome WT (2014) Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Current Biology* Available at: <http://www.sciencedirect.com/science/article/pii/S0960982214006149>.
- Kiani R, Shadlen M (2009) Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. *Science* 324:759–764.
- Krugel L, Biele G, Mohr P, Li S-C, Heekeren H (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences* 106:17951–17956.
- Lau B, Glimcher P (2005) DYNAMIC RESPONSE-BY-RESPONSE MODELS OF MATCHING BEHAVIOR IN RHESUS MONKEYS. *Journal of the Experimental Analysis of Behavior* 84:555–579.
- Law C-T, Gold J (2008) Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat Neurosci* 11:505–513.
- Lee D, Seo H (2007) Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. *Annals of the New York Academy of*

- Sciences Available at:
- <http://onlinelibrary.wiley.com/doi/10.1196/annals.1390.007/full>.
- Lee SH, Dan Y (2012) Neuromodulation of brain states. *Neuron* Available at:
- <http://www.sciencedirect.com/science/article/pii/S0896627312008203>.
- Lee S, Shimojo S, O'Doherty J (2014) Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron* 81.
- Leech, Sharp (2013) The role of the posterior cingulate cortex in cognition and disease. *Brain* 137:1232.
- Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* 453:869–878.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Mansouri F, Tanaka K, Buckley M (2009) Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience* 10:141–152.
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience* 10:647–656.
- McCoy A, Crowley J, Haghigian G, Dean H, Platt M (2003) Saccade Reward Signals in Posterior Cingulate Cortex. *Neuron* 40.
- McCoy A, Platt M (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature Neuroscience* 8:1220–1227.

- McGuire JT, Nassar MR, Gold JI, Kable JW (2014) Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron* 84:870-881.
- Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2003) Dopamine Neurons Can Represent Context-Dependent Prediction Error. *Neuron* 41.
- Nassar M, Rumsey K, Wilson R, Parikh K, Heasly B, Gold J (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci* 15:1040–1046.
- Nassar M, Wilson R, Heasly B, Gold J (2010) An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *Journal of Neuroscience* 30:12366–12378.
- Nieuwenhuis S, Heslenfeld D, Geusau N, Mars R, Holroyd C, Yeung N (2005) Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage* 25.
- O'Reilly J, Jbabdi S, Rushworth M, Behrens T (2013) Brain Systems for Probabilistic and Dynamic Prediction: Computational Specificity and Integration. *PLoS Biology* 11.
- Palomero-Gallagher N, Vogt B, Schleicher A, Mayberg H, Zilles K (2009) Receptor architecture of human cingulate cortex: Evaluation of the four-region neurobiological model. *Human Brain Mapping* 30:2336–2355.
- Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty J (2013) The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. *Neuron* 79.

- Pearce J, Bouton M (2001) THEORIES OF ASSOCIATIVE LEARNING IN ANIMALS. *Psychology* 52:111–139.
- Pearson J, Heilbronner S, Barack D, Hayden B, Platt M (2011) Posterior cingulate cortex: adapting behavior to a changing world. *Trends in cognitive sciences* 15:143–151.
- Pearson J, Platt M (2013) CHANGE DETECTION, MULTIPLE CONTROLLERS, AND DYNAMIC ENVIRONMENTS: INSIGHTS FROM THE BRAIN. *Journal of the Experimental Analysis of Behavior* 99:74–84.
- Porrino L, Goldman-Rakic P (1982) Brainstem innervation of prefrontal and anterior cingulate cortex in the rhesus monkey revealed by retrograde transport of HRP. *Journal of Comparative Neurology* 205:63–76.
- Quilodran R, Rothé M, Procyk E (2007) Behavioral Shifts and Action Valuation in the Anterior Cingulate Cortex. *Neuron* 57.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience* Available at: <http://www.nature.com/nrn/journal/v9/n7/abs/nrn2357.html>.
- Ridderinkhof R, Ullsperger M, Crone E, Nieuwenhuis S (2004) The Role of the Medial Frontal Cortex in Cognitive Control. *Science* 306:443–447.
- Rigotti M, Barak O, Warden M, Wang X-J, Daw N, Miller E, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497:585–590.
- Rushworth M, Behrens T (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience* 11:389–397.

- Rushworth M, Noonan M, Boorman E, Walton M, Behrens T (2011) Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* 70.
- Schacter D, Addis D, Buckner R (2007) Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience* 8:657–661.
- Schultz W, Dayan P, Montague P (1997) A Neural Substrate of Prediction and Reward. *Science* 275:1593–1599.
- Seo H, Lee D (2007) Temporal Filtering of Reward Signals in the Dorsal Anterior Cingulate Cortex during a Mixed-Strategy Game. *The Journal of Neuroscience* 27:8366–8377.
- Shannon CE (2001) A mathematical theory of communication. ... SIGMOBILE Mobile Computing and Communications ... Available at: <http://dl.acm.org/citation.cfm?id=584093>.
- Shenhav A, Botvinick M, Cohen J (2013) The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron* 79.
- Sheth S, Mian M, Patel S, Asaad W, Williams Z, Dougherty D, Bush G, Eskandar E (2012) Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* 488:218–221.
- Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* Available at: <http://www.sciencemag.org/content/296/5573/1709.short>.
- Silvetti M, Alexander W, Verguts T, Brown J (2014) From conflict management to reward-based decision making: Actors and critics in primate medial frontal

- cortex. *Neuroscience & Biobehavioral Reviews* 46:4457.
- Silvetti M, Seurinck R, Verguts T (2011) Value and Prediction Error in Medial Frontal Cortex: Integrating the Single-Unit and Systems Levels of Analysis. *Frontiers in Human Neuroscience* 5.
- Stern E, Gonzalez R, Welsh R, Taylor S (2010) Updating Beliefs for a Decision: Neural Correlates of Uncertainty and Underconfidence. *The Journal of Neuroscience* 30:8032–8041.
- Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *science* Available at: <http://www.sciencemag.org/content/304/5678/1782.short>.
- Sutton RS, Barto AG (1998) Introduction to reinforcement learning. 135 Available at: <http://www.cs.utexas.edu/sites/default/files/research/documents/1%20intro%20up%20to%20RL%3ATD.pdf>.
- Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–1645 Available at: <http://www.sciencemag.org/content/307/5715/1642.short>.
- Vogt B, Gabriel M (1993) Neurobiology of cingulate cortex and limbic thalamus. Available at: <http://link.springer.com/book/10.1007/978-1-4899-6704-6>.
- Vogt B, Vogt L, Laureys S (2006) Cytology and functionally correlated circuits of human posterior cingulate areas. *NeuroImage* 29:452466.
- Vogt BA, Pandya DN (1987) Cingulate cortex of the rhesus monkey: II. Cortical

- afferents. *J Comp Neurol* 262:271–289.
- Vogt BA, Rosene DL, Pandya DN (1979) Thalamic and cortical afferents differentiate anterior from posterior cingulate cortex in the monkey. *Science* 204:205–207.
- Walton M, Croxson P, Behrens T, Kennerley S, Rushworth M (2007) Adaptive decision making and value in the anterior cingulate cortex. *NeuroImage* 36 Suppl 2:T142–54.
- Wilson RC, Nassar MR, Gold JI (2013) A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Comput Biol* 9:e1003150.
- Yu A, Dayan P (2004) Uncertainty, Neuromodulation, and Attention. *Neuron* 46.