Neural signatures of experience-based improvements in deterministic decision-making

Joshua J. Tremel^{a*}, Patryk A. Laurent^b, David A. Wolk^c, Mark E. Wheeler^d, and Julie A. Fiez^a

^aUniversity of Pittsburgh, Pittsburgh, PA, USA. ^bJohns Hopkins University, Baltimore, MD, USA. ^cUniversity of Pennsylvania, Philadelphia, PA, USA. ^dGeorgia Institute of Technology, Atlanta, GA, USA.

*Address correspondence to: Joshua J. Tremel (j.tremel@pitt.edu), University of Pittsburgh, 615 Learning Research and Development Center, 3939 O'hara Street, Pittsburgh, PA 15260.

Abstract

Feedback about our choices is a crucial part of how we gather information and learn from our environment. It provides key information about decision experiences that can be used to optimize future choices. However, our understanding of the processes through which feedback translates into improved decision-making is lacking. Using neuroimaging (fMRI) and cognitive models of decision-making and learning, we examined the influence of feedback on multiple aspects of decision processes across learning. Subjects learned correct choices to a set of 50 word pairs across eight repetitions of a concurrent discrimination task. Behavioral measures were then analyzed with both a drift-diffusion model and a reinforcement learning model. Parameter values from each were then used as fMRI regressors to identify regions whose activity fluctuates with specific cognitive processes described by the models. The patterns of intersecting neural effects across models support two main inferences about the influence of feedback on decision-making. First, frontal, anterior insular, fusiform, and caudate nucleus regions behave like performance monitors, reflecting errors in performance predictions that signal the need for changes in control over decision-making. Second, temporoparietal, supplementary motor, and putamen regions behave like mnemonic storage sites, reflecting differences in learned item values that inform optimal decision choices. As information about optimal choices is accrued, these neural systems dynamically adjust, likely shifting the burden of decision processing from controlled performance monitoring to bottom-up, stimulus-driven choice selection. Collectively, the results provide a detailed perspective on the fundamental ability to use past experiences to improve future decisions.

Keywords: memory; feedback; reinforcement learning; drift-diffusion model; basal ganglia; control.

1. Introduction

Decision-making is fundamentally influenced by past experiences. On a daily basis, we face numerous decisions requiring us to draw upon learned information to optimize our choice selection. This underlying information is acquired through experience, often by using feedback to assess the quality of decisions. Feedback thus provides a critical link between decision outcomes and mnemonic information that influences future choices. However, the processes and neural substrates that translate feedback into improved decision-making remain poorly understood.

The present study investigated how the neural systems that process feedback can influence subsequent decision-making behavior. Specifically, we used computational models of learning and choice selection to investigate how feedback influences decision-making in the context of repeated choice experiences with deterministic outcomes. In such contexts, feedback from a single choice can provide sufficient information to optimize future decisions involving the same choice. For example, if a student guesses that Harrisburg is the capital of Pennsylvania and is then told this answer is correct, the student gains valuable information that may help for later choices. In practice, however, the retrieval of learned information can be faulty, so repetition is typically required to attain optimum performance. These considerations, wherein outcome feedback and choice selection are conceptually linked, suggest that feedback processing and decision-making share a common set of neural substrates. We sought to characterize these substrates using functional magnetic resonance imaging (fMRI) combined with model-informed data analysis.

This work is guided by a literature on perceptual decisions and the impact of noisy sensory input, such as in tasks wherein subjects decide whether a noisy image depicts a face or a

1
2
3
4 house (Heekeren et al., 2004; Dunovan et al., 2014; Tremel and Wheeler, 2015) or whether a
5
6 cloud of dots with partial motion coherence is drifting leftward or rightward (Shadlen and
7
8 Newsome, 2001). These perceptual decision processes are well characterized by sequential
9
10 sampling models, such as the drift-diffusion model (Ratcliff, 1978; Ratcliff and McKoon 2008).
11
12 In this class of model, a decision-variable accumulates evidence for or against a choice—when
13
14 this value reaches a threshold, a decision is executed. As one makes a perceptual choice,
15
16 competing sensory evidence is evaluated until a stopping criterion is reached. This process has
17
18 been conceptually mapped onto physiological measures, wherein the firing rate of neurons
19
20 modulates in a manner analogous to parameters within drift-diffusion models (Hanes and Schall,
21
22 1999; Shadlen and Newsome, 2001; Ratcliff et al., 2007; Roitman and Shadlen, 2002; Glimcher,
23
24 2003). Parallel neuroimaging work in humans has identified fMRI signatures of comparable
25
26 processes (Ploran et al., 2007, 2011; Wheeler et al., 2008; Kayser et al., 2010a, 2010b) and
27
28 connected these signatures directly to drift-diffusion parameters (Bowman et al., 2012; Tremel
29
30 and Wheeler, 2015). Thus, the processes encapsulated by drift-diffusion models are neurally
31
32 plausible and carry specific cognitive interpretations for underlying neural correlates based on
33
34 past work.
35
36
37
38
39
40
41
42

43 Importantly, under the drift-diffusion framework, the source of information to be
44
45 evaluated for a decision can derive from any of a number of systems (e.g., sensory, memory,
46
47 etc.). Since the underlying information being evaluated is the the distinguishing characteristic of
48
49 mnemonic versus perceptual decisions, the predictions of drift-diffusion models should hold for
50
51 decisions wherein evidence is based on information gathered from experience. To support this
52
53 extension, Yang and Shadlen (2007) trained monkeys to associate shapes with reward
54
55 probabilities, and using a modified weather prediction task, linked changes in the firing rates of
56
57
58
59
60
61
62
63
64
65

1
2
3
4 parietal neurons to the integration of probabilistic decision evidence acquired from training. In
5
6 humans, Wheeler and colleagues (2014) demonstrated similar modulations of fMRI activity in
7
8 occipital, temporal, and parietal regions that corresponded to sequences of presented
9
10 probabilistic evidence. In both cases, decision evidence derived directly from arbitrary stimulus-
11
12 response associations acquired through learning, supporting the link between drift-diffusion
13
14 principles and decision-making in non-perceptual domains. On top of this, Frank et al. (2015)
15
16 offered a direct connection by showing that changes in estimates of outcome reward (via
17
18 probabilistic reinforcement learning) correlated with changes in model-estimated decision
19
20 thresholds and that this relationship could be traced to activity in thalamic and medial frontal
21
22 regions. The study's approach, however, focused on specific cortico-striatal connections and did
23
24 not directly consider the broader influence of feedback on decision processes or a feedback-
25
26 based learning framework. Taken together, key findings from simple perceptual decisions seem
27
28 to translate to the mnemonic domain, but the larger intersection between memory and decision-
29
30 making remains unclear.
31
32
33
34
35
36
37

38 One fundamental unknown is how feedback influences decision-making processes and
39
40 improves later behavior. This is especially applicable for deterministic or quasi-deterministic
41
42 decisions, where the available choices always or nearly always produce the same outcome.
43
44 Tricomi and Fiez (2008, 2012) made progress on this front by examining decision-making across
45
46 multiple rounds of a paired-associate learning task. They found that feedback influenced both
47
48 explicit and implicit memory, which together seemed to drive learning and subsequent decision-
49
50 making. However, the Tricomi and Fiez work did not consider model-based connections between
51
52 behavioral and neuroimaging data, and therefore could not draw inferences in the context of
53
54 computationally defined feedback and decision-making processes.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 The current study takes this next step by investigating the influence of feedback on
5
6 decision processes during a concurrent discrimination learning task. In this task, subjects are
7
8 presented with a pair of items, wherein one is arbitrarily designated as the correct choice, and the
9
10 other as incorrect. After selecting an item, subjects see feedback that indicates the value of the
11
12 choice (i.e., a correct or incorrect choice). With repeated experience of the items, subjects accrue
13
14 mnemonic information and gain proficiency in making correct choices. At the same time,
15
16 subjects could also accrue all of the mnemonic information necessary for perfect performance
17
18 from a single decision experience. Thus, concurrent discrimination offers a meaningful
19
20 framework with which to study how the encoding and accrual of past experiences translate into
21
22 improved decision-making.
23
24
25
26
27

28 To capture these feedback-based learning influences, we implemented a reinforcement
29
30 learning model alongside the drift-diffusion model. Reinforcement learning describes how a
31
32 history of outcomes, built up for individual choices, creates a prediction of an outcome's value
33
34 that can be used to drive choice selection. Errors between the expected versus observed
35
36 outcomes (reward prediction error, RPE) are used to adjust the expected value of a choice (EV)
37
38 toward a more accurate representation of the actual result. Thus, measures derived from
39
40 reinforcement learning can explain aspects of feedback processing in terms of a learning signal
41
42 (RPE) and the consequence of that learning (EV). Neurally, these signals have been shown to be
43
44 distinct and separable: in the context of probabilistic learning, each signal tends to localize to
45
46 different areas, such as sub-regions of the basal ganglia and orbitofrontal cortex (Schultz et al.,
47
48 1998, 2000). In the case of deterministic learning, such as concurrent discrimination,
49
50 reinforcement learning should be equally applicable. In an extreme case, where mnemonic
51
52 encoding and retrieval processes are entirely noise-free, the RPE signal from an initial choice
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 should drive sufficient learning to ensure optimal choice selection for a future episode.
5
6 Reinforcement learning can capture this type of single-trial learning, such that the initial RPE
7
8 updates the EV to ensure future accuracy. While deterministic learning can be plausibly captured
9
10 by reinforcement learning models, the extent to which estimates of RPE and EV exhibit the same
11
12 neural localization and dissociations compared to probabilistic choice learning remains an open
13
14 question. Regardless, like the drift-diffusion model, reinforcement learning has a set of distinct
15
16 cognitive predictions that correspond to a set of distinct neural correlates.
17
18
19

20
21 Thus, the two models each describe aspects of learning and decision-making that, when
22
23 combined, could offer unique leverage for understanding how feedback influences decision-
24
25 making behavior, and how this influence is implemented neurally. The drift-diffusion model
26
27 describes how information is evaluated to drive decision-making, whereas reinforcement
28
29 learning describes how feedback shapes that information. As such, each model can distinguish
30
31 between performance-related effects (e.g., boundary in the diffusion model and RPE in
32
33 reinforcement learning) and evidence-related effects (e.g., drift-rate in the diffusion model and
34
35 EV in reinforcement learning). When applied to neural data, these two models can be leveraged
36
37 to identify regions associated with performance and evidence monitoring, versus regions that
38
39 provide evidence to decision-making processes.
40
41
42
43
44

45 To investigate these relationships, we acquired behavioral and functional neuroimaging
46
47 (fMRI) data during a concurrent discrimination task. Drift-diffusion and reinforcement learning
48
49 models were fit to the learning behavior to describe separable aspects of decision-making and
50
51 feedback processing. We then localized brain regions whose activity tracked with changes in the
52
53 model parameters and tested for regional overlap in sensitivity to parameter changes across the
54
55
56
57
58
59
60
61
62
63
64
65

two models. Overall, our goal was to meaningfully connect the neural substrates and cognitive processes related to the processing of outcome feedback with those related to decision-making.

2. Materials and Methods

2.1 Subjects

Twenty healthy, right-handed, fluent English speakers participated in a 1-hour behavioral and functional MRI session. Four subjects were excluded for excessive movement during scanning ($N = 3$) or incomplete data due to technical problems ($N = 1$), leaving 16 (12 female) subjects contributing to the final analyses. These subjects aged 19-31 years (mean 22.1) and had normal or corrected-to-normal vision. Informed consent was obtained from all subjects according to procedures approved by the University of Pittsburgh Institutional Review Board. Subjects were compensated \$75 for their time.

2.2 Task

Subjects performed a 50-item concurrent discrimination learning task during the scanning session (Fig. 1a). In each trial, a pair of words was presented in the center of a screen in a single column for three seconds. In each of these pairs, one word was designated as the correct (positive) choice and the other as the incorrect (negative) choice. During the display period, subjects selected one of the words, indicating their selection with a button press, and received feedback about their decision. Feedback was displayed for 1.5 s and indicated either a correct choice (green checkmarks) or an error (red Xs). Trials were separated by a variable inter-stimulus interval of 1.5-7.5 s selected from a distribution positively skewed toward the shorter intervals (1.5 s increments, mean trial separation = 2.91). This served as jitter between trials to facilitate later deconvolution of the fMRI signal (Dale, 1999). After all 50 word pairs in the set

1
2
3
4 were presented, subjects were allowed to rest for one minute before beginning the next Round, in
5
6 which subjects saw the same set of word pairs again in a randomized presentation sequence and
7
8 again made choices for each pair. Over the course of eight Rounds (in which each discrimination
9
10 was presented once), subjects learned which item in each pair was associated with positive
11
12 feedback (i.e., was “correct”) and which was associated with negative feedback.
13
14

15
16 The discrimination learning task was presented using E-prime and projected onto a screen
17
18 at the head of the magnet bore using a BrianLogics MRI Digital Projection System. Subjects
19
20 viewed the screen via a mirror attached to the radio frequency coil and indicated their responses
21
22 via a fiber optic response glove on their right hand connected to a desktop computer with a serial
23
24 response box (Brain Logics, Psychology Software Tools, Pittsburgh, PA). Earplugs were
25
26 provided to minimize discomfort due to scanner noise.
27
28
29

30 31 *2.3 Stimuli* 32

33 Word stimuli were pulled from the MRC Psycholinguistic Database (Coltheart, 1981).
34
35 All words were one syllable, three to five letters, three or four phonemes, and had a log HAL
36
37 frequency of at least 7.0. Stimuli had concreteness, imageability, and familiarity measures of at
38
39 least 400. Additional word frequency data were obtained from the English Lexicon Project
40
41 (Balota et al., 2007). For each subject, a list was generated by randomly selecting two pre-made
42
43 lists from a set of eight 50-word lists balanced and matched for frequency and number of letters.
44
45 These two lists were individually randomized and then paired, ensuring that each subject had a
46
47 unique set of word pairs balanced for selected psycholinguistic criteria.
48
49
50
51

52 53 *2.4 Image Acquisition* 54

55 Images were acquired on a Siemens Allegra 3-Tesla system. T1-weighted anatomical
56
57 images were obtained using an MP-RAGE sequence (repetition time, TR = 1540 ms; echo time,
58
59
60
61
62
63
64
65

TE = 3.04 ms; flip angle, FA = 8 degrees; inversion time, TI = 800 ms; 1 mm isotropic voxels acquired in 192 sagittal slices). T2-weighted anatomical images were obtained with a spin-echo sequence (TR = 6000 ms, TE = 73 ms, FA = 150 degrees, 0.78 mm² in-plane resolution; 38 axial slices spaced 3.2 mm apart). Functional images sensitive to the BOLD contrast were acquired using a whole-brain echo-planar T2*-weighted series (TR = 1500 ms, TE = 25 ms, FA = 60 degrees, 3.125 mm² in-plane resolution, 29 axial slices spaced 3.5 mm apart). The first four images of each run were discarded to allow for net magnetization and RF equilibration.

2.5 Functional Imaging Preprocessing

Image preprocessing and analysis was carried out using FIDL, a software package developed at Washington University in St. Louis. Imaging data were preprocessed to address noise and image artifacts, including within-TR slice-time acquisition correction, motion correction using a rigid-body translation and rotation algorithm (Snyder, 1996), within-run voxel intensity normalization of 1000 to facilitate cross-subject comparisons (Ojemann et al., 1997), and computation of a Talairach atlas space transformation matrix (Talairach and Tournoux, 1988). Single-subject analyses were performed in data space. For group analysis, data were resampled to 2 mm isotropic voxels and transformed into Talairach atlas space.

2.6 Behavioral analysis

Behavioral accuracy and response time (RT) measures were assessed across Rounds (repetitions of the 50-item set) using repeated-measures ANOVAs. Round was entered as an eight-level (Round 1-8) within-subjects factor. For all ANOVA analyses, the Greenhouse-Geiser sphericity correction was applied when Mauchly's test indicated a violation of the sphericity assumption.

2.7 Drift-diffusion modeling

1
2
3
4 The drift-diffusion model is a robust model of decision-making behavior that describes
5
6 variations in behavioral response time distributions on the basis of several parameters. Of interest
7
8 for this study were mean drift-rate (v), corresponding to the mean rate of evidence accumulation
9
10 for a condition, and decision threshold (a), corresponding to decision criteria or the amount of
11
12 evidence needed before committing to a decision (Ratcliff, 1978; Ratcliff and McKoon, 2008).
13
14 We expected that as subjects learned, their decision-making would become more proficient. In a
15
16 drift-diffusion model, an increase in the drift-rate and decrease in the threshold parameters would
17
18 correspond to faster and more confident decisions. Changes in behavior, as captured by changes
19
20 in the model parameters, should thus map onto changes in the activity of brain regions associated
21
22 with evidence accumulation (modeled by drift rate parameter) and decision criteria (modeled by
23
24 the threshold parameter). It is important to note that because regional activity might correlate
25
26 with changes in a parameter does not necessarily mean that particular region is responsible for
27
28 the computation itself (e.g., if a region correlates with changes in drift-rate, this region does not
29
30 necessarily function as an evidence accumulator). Rather, the correlation is a marker of the
31
32 cognitive process—the region is associated in some way with what the parameter represents. For
33
34 instance, a region whose activity correlates with drift-rate may not be responsible for computing
35
36 a decision variable itself, but rather is associated with that process, either responding to other
37
38 regions computing the variable or providing signals that feed into the computation. Regardless,
39
40 however, the region would be associated with drift-rate, while not necessarily computing it.
41
42
43
44
45
46
47
48
49

50 Drift-diffusion models were fit to subject reaction time and choice data using the HDDM
51
52 software package (Wiecki et al., 2013). HDDM fits models using hierarchical Bayesian methods
53
54 wherein subject-level and group-level parameters are estimated at the same time using Markov-
55
56 chain Monte-Carlo (MCMC). To examine changes in key parameter estimates across learning,
57
58
59
60
61
62
63
64
65

two parameters, mean drift rate (v) and decision threshold (a), were allowed to vary across the eight Rounds of the task (i.e., eight repetitions of the 50-item discrimination set). Thus, there was one value per Round for each of drift-rate and threshold, allowing the model to fit changes in behavior across learning in terms of these two parameters. The non-decision time (t_{er}) and starting point (z) were estimated at the subject- and group-levels, while variance parameters (variance in drift, sv , non-decision time, st , and starting point, sz) were estimated only at the group level. Data were accuracy-coded, such that the upper threshold (a) of the model corresponded to a correct choice (i.e., the subject chose the positive item), whereas the lower bound (0) corresponded to errors (i.e., the subject chose the negative item). Round-wise parameter estimates for drift-rate and threshold were used as covariates in the imaging analyses.

Model parameters were estimated using three MCMC chains of 10,000 samples each. The first 1,000 samples of each chain were discarded to ensure that the chain stabilized. After sampling, model convergence was evaluated visually by inspecting the traces of the model posteriors and statistically by computing the Gelman-Rubin statistic, which compares inter- and intra-chain variance with an ANOVA approach. This statistic was near 1.0 (maximum deviance from 1.0 was 0.007) for all parameters, suggesting that the models converged properly. For a third check, the Geweke statistic, which compares variance at the starts and ends of chains, also indicated proper convergence.

Model fits to the data were evaluated by simulating choice and reaction time data from the model posterior distributions and comparing to the observed data. One hundred datasets were generated from each subject's model. The mean of these 100 sets was then used to compare against the empirical data to ensure that model-derived data can reproduce key patterns of accuracy and response time in the observed behavioral data. Figure 1b illustrates the relationship

of the model-simulated versus empirical behavior. These were assessed for differences across conditions using 2 x 8 repeated measures ANOVA including factors of Dataset (observed, model-derived) and Round (Round 1-8).

Two alternative models were also constructed to assess whether the aforementioned model indeed provided the best fit to the data. One alternative model allowed only threshold to vary by Round, while the other allowed only drift-rate to vary by Round. Primary and alternative models were assessed with the deviance information criterion (DIC), which measures the lack of fit of model estimates and penalizes for complexity (i.e., degrees of freedom) (Spiegelhalter et al., 2002; Spiegelhalter et al., 2014). A lower DIC indicates a better fit, typically by a magnitude of 10 or more (Burnham and Anderson, 2004; Zhang and Rowe, 2014).

2.8 Reinforcement learning agent

In order to capture the dynamics of feedback-based experience across Rounds as subjects performed the concurrent discrimination learning task, a reinforcement learning agent was used. This agent used an off-policy temporal difference control algorithm (Q-learning) to estimate reward-prediction error (RPE) signals and item-level expected values (EV) from subjects' choice selection behavior across learning (Watkins and Dayan, 1992; Sutton and Barto, 1998). The agent computed an EV (quality measure or Q-value) for selecting each of two actions in a given state. Here, a state was defined as a word pair. A subject, then, could take two possible actions—choosing one or the other word in the pair. The agent updated the EV based on the human subjects' choices by incrementing the value with an RPE (multiplied by a constant learning rate). EVs of each action were instantiated with random values ranging from 0 to 1 selected from a uniform distribution. Through learning, these initial values should approach the observed value for that action. The RPE was calculated as the difference between the observed outcome (1 for a

correct response, 0 for an error) versus the EV for a given action. Thus, the agent updated its values based on the experiences of the human subject.

In contrast to the drift-diffusion model, the reinforcement learning agent learned at the trial-level and learned only from the choice data (i.e., did not address variance in response times). Thus, RPE and EV estimates derived from the model were trial-by-trial estimates and not round-level averages (cf. threshold and drift-rate in the drift-diffusion model). EVs were entered into a softmax logistic function to compute response probabilities for the chosen action. From these probabilities, we computed a log likelihood estimate reflecting the likelihood that a given subject would respond a particular way across the task session. An agent was considered to be optimal when its learning rate parameter produced the maximum likelihood. The models and optimization routine were implemented within Matlab (The MathWorks, Natick, MA). The input learning rate parameter was sampled across a range of possible values from 0.0 to 1.0. Trial-wise model estimates of RPE and EV were extracted from the model running with the optimal learning rate. These estimates were used as covariates in the imaging analyses. The extent of model fit was measured with pseudo- R^2 statistics computed as $(LL_{\text{model}} - LL_{\text{null}})/LL_{\text{null}}$, wherein LL_{model} is the log-likelihood estimate of the fitted agent and LL_{null} is that of a chance-performance model (i.e., response probabilities are 0.5 for all trials). These statistics are presented in Table 2.

2.9 Voxelwise imaging analysis

Subject-level neuroimaging data were analyzed using a voxel-by-voxel general linear model (GLM) approach. In the GLM, each time point is computed as the sum of coded effects, produced by model events and by error (Friston et al., 1994; Glover, 1999; Miezin et al., 2000; Ollinger et al., 2001). The GLM approach makes no assumptions about the underlying shape of

1
2
3
4 the blood-oxygen-level-dependent (BOLD) response, but assumes that component signals,
5
6 described by the coded events, will sum linearly to produce the observed BOLD signal. Within
7
8 each Round, a linear term was included to capture signal drift, and a constant term captured
9
10 baseline signal. A high-pass filter at 0.009 Hz was added to the models to reduce the influence of
11
12 low-frequency fluctuations in the BOLD response. Each trial was modeled with a series of finite
13
14 impulse response (FIR) basis functions, wherein each time point of a trial was described by a
15
16 delta function. The sequence of these functions corresponded to trial-level time series averages.
17
18 Event-related effects were described as an estimate of the percent of BOLD signal change
19
20 relative to the baseline constant.
21
22
23
24
25

26 For each subject, two GLMs were created which separately modeled effects related to
27
28 diffusion model parameters and effects related to reinforcement learning measures. In both
29
30 GLMs, accuracy was entered as a categorical event-related regressor, which coded each decision
31
32 trial as correct or incorrect, modeled to 11 time-points (16.5 s). In the diffusion model GLM, two
33
34 additional covariates were entered to describe changes in the drift-rate and decision threshold
35
36 parameters across Rounds. These were modeled as continuous measures across the entire task
37
38 and thus captured correlation between the modeled measure and the underlying BOLD signal.
39
40 On a trial-level, these covariates were modeled with 11 FIR basis functions (11 time-points
41
42 totaling 16.5 seconds) to describe changes in correlation within a trial epoch. Importantly, the
43
44 diffusion model parameters captured choice and response time behavior on the level of Round
45
46 (i.e., all trials in Round 1 had the same parameter values, all trials in Round 2 had the same
47
48 parameter values, etc.) (Table 1). Modulations in these parameters therefore reflected
49
50 modulations by Round.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

The other GLM modeled covariates derived from the reinforcement learning models. RPE and EV measures were entered as covariates to model trial-by-trial modulations across the entire task in a manner similar to the diffusion model GLM. In contrast to the diffusion model parameters (modeled Round-by-Round), the positive RPE and EV measures were continuous such that each trial had a unique value (modeled trial-by-trial). Round was not coded in either GLM since the covariate measures derived from both computational models capture variances due to Round (i.e., measures inherently modulate across Rounds, thus changes in the measures reflect changes across Round).

Importantly, only positive RPEs were entered. There are several reasons for this analysis choice. The first is theoretical. Specifically, positive RPEs are thought to capture learning signals derived from feedback and correlate with phasic increases in the firing rates of dopamine neurons in the brainstem and striatum (Schultz et al., 1997; O'Doherty et al., 2003; Bayer and Glimcher, 2005; Daw and Doya, 2006; Oyama et al., 2010). In contrast, negative prediction errors seem to correspond to different processes than positive ones and localize to different regions in the brain (Spoormaker et al., 2011). If negative RPEs were included in the analysis, it could potentially complicate interpretations and under-power results related to the more dominant positive prediction errors that seem to drive most of the learning in this task (i.e., negative prediction errors are relatively rare in this task, especially in late rounds). Second, very few of the total trials were associated with negative RPEs. Across subjects, a median of 8% of the total trials (~32 trials) were negative RPE trials. Given the single-subject regression design of the GLMs, this is considered to be too few trials to get reliable estimates from a single regressor (Green, 1991). Indeed, the observed power for detecting a moderate effect size for a linear multiple regression or ANCOVA with this design and number of trials would be about 0.20. In

1
2
3
4 contrast, the observed power for the same analysis for the positive RPE regressor would be at
5
6 least 0.95 for all subjects. Thus, while negative RPE signals have clear theoretical value, this
7
8 phenomenon was not adequately sampled in the present study.
9

10
11 A critical concern about this regression-style approach to the imaging analysis is that the
12
13 covariates between the two GLMs are not necessarily orthogonal. Indeed, the correlations
14
15 between the four covariates of interest are generally high on a subject level, meaning the models
16
17 will attempt to parse variance based on the uncorrelated components. This suggests that
18
19 separating the effects attributable to each of the covariates is difficult and should be considered
20
21 with a degree of caution (for example, Hare et al., 2008; Hunt, 2008). While these concerns
22
23 should be kept in mind, there are a few factors that may partly assuage these issues. First, since
24
25 the covariates of interest are derived from computational models that offer strong predictions
26
27 regarding neural substrates, results from each covariate (Figs. 2 and 3) can be compared against
28
29 the prior literature. Replication of prior findings bolsters confidence in the present findings.
30
31 Second, between the two models, our analysis seeks to associate parameters between the models
32
33 rather than parse them (e.g., examining overlap of regions that track drift-rate with those that
34
35 track EV). In contrast to much other work, our objective is not to explain effects solely in terms
36
37 of decision-making or reinforcement learning, but rather in terms of associations between the
38
39 models to gain insight into the neural substrates that are shared between decision-making and
40
41 learning. Establishing such associations between these two models is an important step in
42
43 identifying possible neural substrates that mediate the influence of feedback on decision-making.
44
45
46
47
48
49
50
51

52 *2.10 Separating neural effects of within-model parameters*

53
54

55 Since the measures within each model are related both behaviorally and conceptually, our
56
57 focus was on regions that predominantly track one or the other within-model measure (e.g., RPE
58
59
60
61
62
63
64
65

more than EV or vice-versa). To statistically separate the effects modeled within each GLM, voxel-by-voxel paired t-tests were computed to identify voxels that differentially correlated with the parameters from the diffusion model, or with the parameters from the reinforcement learning agent. For the diffusion model GLM, the drift-rate covariate was tested against the decision threshold covariate. Voxels in the resulting image with a positive t-statistic indicated underlying activity that correlated to a significantly greater degree with modulations in decision threshold than drift-rate. In other words, voxels with statistically larger correlation coefficients for threshold were indicated by significantly positive t-statistics in the activation map. Likewise, voxels with a negative t-statistic indicated activity that correlated to a significantly greater degree with modulations of drift-rate than decision threshold. As such, this t-test separated activity that was statistically more attributable to modulations in either parameter, as computed by the linear models. The analysis was restricted to time points 4-7 (of the 11 modeled time-points), which captured the bulk of trial-related activity and excluded biologically irrelevant time points (e.g., first and last time points of the series). An identical approach was taken with the reinforcement learning measures for the other GLM, separating activity that was statistically more attributable to modulations in either RPE or EV.

Images resulting from the t-tests were corrected for multiple-comparisons using a cluster-extent threshold and adjusted for sphericity (minimum z-transformed t-statistic of 2.5, $p = 0.01$, 42-voxel contiguity). These corrected images were then smoothed with a two-voxel (4 mm) full width at half maximum (FWHM) Gaussian kernel. A threshold was applied to the smoothed, corrected images at a Z-statistic of 2.0 to eliminate any sub-threshold bleed-over resulting from the smoothing procedure (i.e., from the tails of the Gaussian kernel).

1
2
3
4 These maps were used to define regions of interest whose activity correlated with the
5
6 individual covariates (e.g., regions whose activity correlated with modulations in positive RPE
7
8 signals). Z-statistics that exceeded an alpha of 0.001 were tagged as peaks, around which
9
10 spherical regions (12 mm radius) were grown. Voxels within the spheres that were absent from
11
12 the corrected maps were then excluded from the region (i.e., voxels that did not pass the multiple
13
14 comparisons and sphericity tests). The coordinates of the peaks of each of the resulting regions
15
16 were used to identify approximate anatomic locations in Talairach space (Talairach and
17
18 Tournoux, 1988). Regions were defined for each covariate, resulting in a set of regions of
19
20 interest that corresponded to effects related to drift-rate, decision threshold, RPE, and EV. The
21
22 purpose of this region identification procedure was to examine regions related to each covariate
23
24 and to relate it to previous literature. Thus, this level of analysis was independent of the
25
26 following intersection analysis.
27
28
29
30
31
32

33 *2.11 Intersection of model effects*

34
35

36 To establish meaningful links between the two computational approaches and the neural
37
38 signatures of each, an analysis of the intersection of significant effects, similar to multi-
39
40 experiment conjunctions, was performed. This analysis identified regions corresponding to the
41
42 functional overlap between diffusion model and reinforcement learning parameters. First, for
43
44 each measure of interest (RPE, EV, decision threshold, and drift-rate), a binary mask was created
45
46 that marked any above-threshold (i.e., significant) voxel with a 1 and any below-threshold voxel
47
48 with a 0. These masks were computed from the corrected statistical images that resulted from the
49
50 t-tests described above. Thus, each mask contained information identifying which voxels
51
52 correlated with each of the four measures at a relatively rigorous threshold corrected for multiple
53
54 comparisons and adjusted for sphericity.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 The resulting masks for the diffusion model parameters were then compared to the masks
5
6 of the reinforcement learning measures. There were, thus, four possible combinations: Threshold
7
8 + RPE, Threshold + EV, Drift-rate + RPE, and Drift-rate + EV. For each combination, a logical
9
10 “AND” computation identified voxels that exhibited above-threshold effects in both source
11
12 masks (e.g., a given voxel’s activity correlated significantly with threshold modulations and
13
14 correlated significantly with RPE modulations in separate analyses). This computation is
15
16 analogous to the minimum t-field computation (or minimum statistic) of multi-experiment
17
18 conjunctions, which ensures that a given voxel shows a reliable effect independently for both
19
20 covariate and not just for one. The null hypothesis for this statistic is that a voxel’s activity
21
22 across subjects correlated with at most, one of the covariates. Therefore, rejecting the null
23
24 hypothesis would mean that a voxel’s activity correlated with both covariates. For each
25
26 combination, the result was an image of voxel clusters that were functionally correlated with
27
28 both model parameters. This analysis, then, related functionally identified regions of
29
30 reinforcement learning to regions associated with specific aspects of a decision-making
31
32 processes. We computed the center of mass for each region in the intersections to identify
33
34 approximate anatomical locations in Talairach space (Talairach and Tournoux, 1988).
35
36
37
38
39
40
41
42

43 The choice of threshold for this analysis is, to an extent, arbitrary. There is a considerable
44
45 compromise that must be made with respect to false positives and false negatives: by changing
46
47 the threshold, the results could change appreciably. Our approach is reasonably cautious with
48
49 false positives and favors a higher threshold at the risk of potentially excluding overlapping
50
51 regions by mistake. The maps presented in Figs. 2 and 3 show the statistical values of the
52
53 individual covariates that may help assess the extent to which this tradeoff might impact the
54
55 results.
56
57
58
59
60
61
62
63
64
65

3. Results

3.1 Behavioral performance

Subjects overall learned quickly, reaching 80% accuracy in three to four repetitions (Figure 1b). This improvement in accuracy across rounds was indicated by a significant main effect of Round in a repeated measures ANOVA ($F[2.58, 38.65] = 83.35, p < 0.001, \eta_p^2 = 0.85$) and significant rank-order correlation ($\rho = 0.99, p < 0.001$). Likewise, response time (Figure 1c) decreased across rounds beginning at the first recall round (i.e., Round 1 was exposure only and required no retrieval of learned information), indicated by a significant Round effect on response times ($F[2.16, 32.33] = 19.12, p < 0.001, \eta_p^2 = 0.56$) and negative rank order correlation ($\rho = -0.93, p < 0.001$). Altogether, the increase in accuracy across rounds suggests that subjects were able to successfully optimize their choice selections with repeated experience. Likewise, the decrease in response times suggests that decision-making becomes more automatized through learning.

3.2 Drift-diffusion modeling

The primary goal of the diffusion model analysis was to attempt to capture behavioral improvements across rounds and describe those changes in terms of decision-making processes. If these changes can be captured successfully, model parameters could be used to inform neuroimaging analysis and thus examine the neural underpinnings of improvements to decision processes across learning. Of specific interest were the threshold and drift-rate parameters, which were free to vary across Round. The model was able to produce parameters that could accurately describe the behavioral data (Fig. 1b). The mean group parameters for the final model are enumerated in Table 1.

In terms of decision processes, improvements in choice selection behavior across rounds manifested as an overall decrease in decision thresholds coupled with an overall increase in drift-rates across rounds. Figure 1c illustrates decision threshold changes, which increased from Round 1 to Round 2, and then gradually decreased over subsequent Rounds. These variations reached statistical significance, indicated by a main effect of Round ($F[3.26, 48.98] = 10.784$, $p < 0.001$, $\eta_p^2 = 0.42$). The changes in the parameter progressively decreased across Rounds, as indicated by a negative rank order correlation, excluding Round 1 in which decision thresholds had to be formed without prior experience of the experimental context ($\rho = -0.99$, $p < 0.001$). Figure 1c also illustrates changes in drift-rates across Round. Drift-rate increased steadily with repetitions of the item set, indicated by a significant effect of Round ($F[1.71, 25.57] = 81.97$, $p < 0.001$, $\eta_p^2 = 0.84$) and positive rank order correlation ($\rho = 0.99$, $p < 0.001$). Altogether, decreases in threshold and increases in drift-rate are intuitive patterns for a learning task such as this, wherein subjects gain confidence in their choices through repetition and learn to make fast and accurate decisions. Moreover, while the pattern of changes across both parameters is intuitive, it is not requisite; it is equally plausible, for instance, that behavior might be described solely based on changes to one parameter or the other. This was not the case. Two alternative models, wherein either drift-rate or threshold was fixed while the other was free to vary by round, produced poorer fits (DIC 10,325 and 9,324) to the observed data than the primary model (DIC 8,927). It is also worth noting that, while parameters in Round 1 do not reflect experience-related computations, the model seems to account for this, as observed in the increase in threshold and change in sign for drift-rate from Round 1 to Round 2. Conceptually, subjects approach Round 1 with trial-and-error guessing and subsequently switch to a controlled retrieval strategy at Round 2.

To assess the fit of the model, choice and response time data were simulated from the posterior distribution of the final model. Data produced by simulation mapped onto the empirical data with little error for both accuracy (Round x Dataset interaction, $F[2.43, 77.84] = 0.06$, $p = 0.96$, $\eta_p^2 = 0.002$) and RT measures ($F[2.03, 64.86] = 0.12$, $p = 0.89$, $\eta_p^2 = 0.004$). These null ANOVA results indicate that the empirical and simulated datasets were not different and suggest that data generated from the model posteriors could reproduce key behavioral patterns within the empirical dataset, on average (Fig. 1b).

3.3 Reinforcement learning agent

In parallel with the diffusion model analysis, a reinforcement learning approach was taken to characterize changes in choice selection behavior across rounds in terms of feedback processing and mnemonic prediction strength. Accuracy data for the concurrent discrimination task indicated gradual improvements to performance as choices were repeatedly experienced. As such, the reinforcement learning agent should produce trial-level EV estimates that gradually increase across rounds (reflecting better outcome predictions) and RPE estimates that gradually decrease in magnitude and frequency across rounds (reflecting fewer unexpected outcomes).

Overall, the pseudo- R^2 statistic of the reinforcement learning models was 0.37 on average, indicating the models fit the human behavior quite well compared to a chance-performance model (Table 2). Learning rates of the models were generally high, at a mean of 0.81 (0.02), corresponding to the rapid climb of the learning curve (Fig. 1b). This indicates that, according to the models, subjects strategically placed a high importance on new information in learning the choice-outcome values. It is worth noting that, while reinforcement learning is typically applied to tasks with probabilistic choice-outcome relationships (in which learning is typically slow), it was successfully applied here to concurrent discrimination learning, which

1
2
3
4 features deterministic choice-outcome relationships and relatively rapid learning. A key result
5
6 here is in the learning rate for our deterministic task, which is considerably higher than learning
7
8 rates in probabilistic learning tasks (Rieskamp, 2006; Doll et al., 2009; Cavanagh et al., 2010;
9
10 Gläscher et al., 2010). Importantly, this high learning rate determines the relative influence of
11
12 EV and RPE values and how important individual choice experiences are to a particular subject.
13
14

15 16 *3.4 Effects of single model covariates*

17
18 Regions of interest corresponding to activity that correlated with the model parameters
19
20 were identified from the covariate activation maps. Tables 3-6 enumerate regions whose activity
21
22 correlated with modulations in positive RPEs, EV, decision threshold, and drift-rate,
23
24 respectively. The correlation maps are presented in Figures 2 and 3. Briefly, the results run
25
26 parallel to many findings in the literature. Positive RPE signals correlated with activity in regions
27
28 such as the pre-supplementary motor area (pre-SMA), caudate nucleus, cingulate gyrus, and
29
30 dorsolateral prefrontal cortex (dlPFC), among other regions (Table 3) (Gläscher et al., 2010;
31
32 Glimcher, 2011; Daw et al., 2011). EV correlated with activity in regions such as the
33
34 parahippocampal gyrus, orbitofrontal cortex, and the putamen (Table 4) (Frank and Claus, 2006;
35
36 Hare et al., 2008). Decision thresholds varied with activity in cognitive control regions such as
37
38 the pre-SMA, anterior insula, and frontoparietal regions (Table 5) (Mulder et al., 2012; Frank et
39
40 al., 2015). Drift-rates varied with activity in memory regions such as the parahippocampal gyrus
41
42 and putamen, as well as control regions like supplementary motor areas (SMA) (Table 6) (Badre
43
44 et al., 2014).
45
46
47
48
49
50
51

52 53 *3.5 Intersection of model-defined neural effects*

54
55 To link together the cognitive processes captured by the drift-diffusion model and
56
57 reinforcement learning, along with their neural substrates, a conjunction analysis was performed.
58
59
60
61
62
63
64
65

1
2
3
4 The overlap of voxels corresponding to learning-related modulations in drift-diffusion model and
5 reinforcement learning parameters were examined using a 2 x 2 design of evaluation (Threshold,
6 Drift-rate by RPE, EV). Each intersection image contained voxels that were significantly active
7 for both of the two parameters in the intersection (e.g., Threshold and RPE). Importantly, these
8 intersections should not be interpreted as indicating dependence between measures (i.e.,
9 interactions), but rather that present voxels exhibited activity that correlated with both
10 modulations of a diffusion model parameter and also to modulations of a reinforcement learning
11 measure. Moreover, to keep interpretations as clean as possible, the intersections were created by
12 separating within-model effects (i.e., effects due to positive RPE signals were statistically
13 separated from those due to EV; the same applied to drift-rate and threshold effects). This
14 method thus prevented any overlap between parameters of the same model, namely between RPE
15 and EV and between drift-rate and threshold.

16
17 The intersections of Drift-rate + RPE and of Threshold + EV produced no overlapping
18 voxels and were thus excluded from further consideration. The two remaining intersections
19 produced sets of 11 and 5 regions corresponding to intersections of Threshold + RPE and Drift-
20 rate + EV covariates, respectively. The overlap of effects related to positive RPE modulations
21 and to changes in decision threshold localized to 11 regions (Table 7), encompassing territory in
22 the caudate head, anterior insula, medial frontal gyrus (pre-SMA), fusiform gyrus, anterior
23 cingulate cortex, right inferior occipital gyrus, and middle frontal gyrus (dlPFC). The other
24 intersection, corresponding to the overlap of drift-rate and EV effects, identified five regions that
25 clustered in the putamen, inferior parietal lobe, SMA, middle insula, and middle temporal lobe
26 (Table 8).

1
2
3
4 It is important to note that while these overlaps can be interpreted with respect to the
5
6 computational models that provide the covariates, a positive overlap result does not necessarily
7
8 indicate that the contained regions are performing the same function or exhibit shared
9
10 processing. For instance, a given region that exhibits a conjunction effect across two particular
11
12 conditions might exhibit dissociable patterns of activity (Woo et al., 2014) or connectivity
13
14 (Smith et al., 2015) between those conditions. However, because the analysis is testing
15
16 correlation based on results from computational models that attach specific interpretations to the
17
18 parameters, we can speculate how processing related to changes in these parameters might
19
20 overlap conceptually. In this way, these correlations do not necessarily implicate the regions in
21
22 the computation of the parameter itself, but rather link it to the conceptual processes represented
23
24 by the parameter in the models. For instance, RPE signals seem to be computed by brainstem
25
26 dopamine neurons (Schultz et al., 1997) and subsequently broadcast diffusely to other regions
27
28 throughout the brain. Though activity in these other regions frequently correlates with RPE
29
30 signals, it is unlikely that they participate in the computation of RPE. Rather, the correlation
31
32 marks a process that results from or is associated with the RPE signal, not performing of the
33
34 computation itself. It is in this spirit that conceptual interpretations of the parameter-correlated
35
36 overlap regions are discussed in the following section.
37
38
39
40
41
42
43
44
45
46
47

48 **4. Discussion**

49

50 In this study, we used a dual model approach to investigate how different aspects of
51
52 decision-making were influenced by feedback about decision performance and how the neural
53
54 correlates of feedback processing and decision-making relate to each other. Behavioral data from
55
56 a concurrent discrimination task were fit using a drift-diffusion model to characterize decision-
57
58
59
60
61
62
63
64
65

making processes and a reinforcement-learning model to describe the influence of feedback. Parameters from these models were regressed onto neural data to localize brain regions whose activity varied with changes in the parameters. Correlations between modulations in the parameters across Rounds and modulations in fMRI activity served as markers of the processes described by the computational models. The intersection of the regions localized across the two models was then examined: brain regions common between the two models should correspond to putative substrates of feedback influences on decision-making.

The patterns of overlap between regions localized by both drift-diffusion and reinforcement-learning model parameters support two primary inferences. First, frontal, fusiform, anterior cingulate, and caudate nucleus regions behave like performance monitors, seeming to reflect errors in performance predictions that signal the need for changes in decision-making control. Second, temporoparietal, SMA, and putamen regions behave like evidence storage and accumulation sites, seeming to reflect differences in learned item values that inform optimal decision choices. As information about optimal choices is accrued, these neural systems dynamically adjust as decision-making becomes more skilled (i.e., faster and more accurate). Figure 4a provides a graphical depiction of these inferred intersections, which are discussed in detail below.

4.1 Outcome feedback influences decision thresholds via performance monitoring

Feedback about choice outcomes provides direct information about whether or not a given behavior met one's performance expectations. In the concurrent discrimination task, an accurate choice in an early round might signal that a chosen memory retrieval strategy worked surprisingly well, whereas an inaccurate choice in later rounds might indicate that a choice was

made too quickly. More generally, outcome feedback provides critical information about whether control over the decision-making process is appropriately matched to performance expectations. Outcome feedback can therefore indicate that controlled processing for a subsequent decision can be safely reduced without jeopardizing performance, or that it should be increased so that more or better evidence can be gathered to support a given choice option.

Feedback about such performance discrepancies can be conceptualized as differences between predicted versus observed outcomes. These differences are captured by RPE signals in reinforcement learning models. Similarly, changes in controlled processing can be conceptualized as changes in the amount of evidence needed before a choice can be made; these are captured by the threshold parameter in drift-diffusion models. Thus, the neural substrates corresponding to the influence of outcome feedback on performance monitoring can be derived from the intersection of regions localized by the positive RPE and threshold parameters.

Activity in the anterior insula, pre-SMA, dlPFC, anterior cingulate cortex, fusiform and inferior occipital gyri, and dorsal caudate head (Table 7) was sensitive to changes in positive RPE signals and decision thresholds. These regions might therefore be important for performance monitoring and adjustments in controlled processing in response to unexpected outcomes. It is worth noting that for perceptual decisions, similar regions have been found to exhibit sensitivity to outcome information (Wheeler et al., 2008). This suggests that the control exerted by these regions is not exclusive to decisions based on retrieved information or to decisions with explicit feedback, and thus might reflect some degree of domain-general functionality.

Consistent with this interpretation, each of these regions has been more broadly implicated in performance monitoring. For example, the anterior insula has been associated with

the conscious perception of errors (Ullsperger et al., 2010), while the anterior cingulate and dorsolateral prefrontal regions seem to be important in conflict monitoring and cognitive control (Botvinick, 1999; MacDonald et al., 2000). Furthermore, the identification of the dorsal striatum (particularly the dorsal caudate head) is consistent with anatomical models that distinguish between the dorsomedial and dorsolateral striatum, with the former associated with cognitive performance evaluation, goal achievement, and action-reward associations (Tricomi and Fiez, 2008, 2012) and the latter associated with sensorimotor learning and the encoding of implicit reward representations (cf. putamen in the Drift-rate + EV intersection) (Poldrack et al., 2005; Williams and Eskandar, 2006; Seger et al., 2010). Regions in the fusiform and inferior occipital gyrus might reflect an executive influence from prefrontal regions to sensory regions or an attentional biasing on perceptual representations. Additionally, some of the regions are associated with behavioral speed-accuracy tradeoffs, including the pre-SMA, dlPFC, striatum, and perceptual regions in the ventral stream (Bogacz et al., 2010). Thus, the wider literature supports the idea that regions identified from the Threshold + RPE intersection mediate the influence of external performance signals on adjustments to decision-making control.

4.2 Feedback history influences the rate of evidence accumulation

High quality decision evidence is critical to optimal choice selection performance. As noted above, in deterministic decisions, outcome feedback can provide unambiguous information about optimal choices. Experience-based improvements in decision-making should therefore reflect the degree to which this information is encoded into memory and successfully retrieved. In the concurrent discrimination task, mnemonic information deemed relevant to a choice episode can be gathered from various sources and integrated as evidence for or against a given

option. For instance, a subject might rely heavily on contextual or associative information from declarative memory in the medial temporal lobe to make an accurate choice on some trials. For other trials (or for other subjects), implicit value information (i.e., familiarity) might drive choice selection, via procedural memory in the basal ganglia. In most cases, it seems likely that a combination of these influences drives optimal decision-making in deterministic contexts (Tricomi and Fiez, 2008, 2012).

Since EV is an analogue of evidence quality, it can be used to localize the neural substrates of mnemonic information that guide choice selection. In parallel, drift-rate provides a measure of the rate of evidence accumulation during a decision. Because EVs are updated based on feedback at the end of each trial and reflect a history of prior experience, the overlap between effects related to this history described by EVs and those related to drift-rate of the current trial reflects the influence of the accrued experience of feedback from previous trials on evidence accumulation. In this intersection, regions were found in inferior parietal lobe, middle temporal gyrus, middle insular cortex, SMA, and the putamen (Table 8).

At its simplest, the concurrent discrimination task requires a mapping of a particular visual item (here, a word) to a motor output (which can change from trial to trial and Round to Round). The SMA, then, may be involved in creating such a mapping on the fly, along with other regions in this intersection, such as the middle insular cortex, which has been implicated in motor control related to habit learning (Poldrack et al., 1999). Importantly, coactivation of similar collections of regions has been reported in studies examining practice-related sensorimotor coordination, timing of movements, and task repetition (Jantzen et al., 2002; Jantzen et al., 2004; Smith et al., 2004). This suggests that the regions in this intersection might contribute to building strong stimulus-driven response behavior over the course of the task. The

1
2
3
4 role of the dorsolateral striatum (putamen) seems to be particularly important in this type of
5
6 procedural habit and motor learning (Yin et al., 2004, 2006; Tricomi et al., 2009). For instance, if
7
8 an EV on a particular trial has been well-learned, drift-rate will be high, suggesting strong
9
10 functional connections between the representation of the EV and the necessary motor mapping to
11
12 enact the correct response given the orientation of the stimuli for that trial. Taken together, this
13
14 suggests that corticostriatal circuits similar to those implicated in procedural memory or habit
15
16 learning might partly underlie performance on a concurrent discrimination task. Such a system
17
18 supported by the putamen would be well suited to contribute salient value information about
19
20 each choice option. This information could be built up over repeated outcome experiences,
21
22 establishing a sense of item-level familiarity (Poldrack et al., 1999; Poldrack et al., 2001; Seger
23
24 and Cinotta, 2005). EV, thus, carries information about a particular stimulus, such that stronger
25
26 EVs can lead to easier choices (i.e., better discriminability). Drift-rate can therefore be directly
27
28 influenced by the quality of an EV, such that stronger EVs can speed up a decision process by
29
30 providing higher quality evidence. A procedural memory system could work to build strong EVs
31
32 to be used as evidence for a choice.
33
34
35
36
37
38
39
40
41
42

43 *4.3 Repeated feedback encourages a shift toward stimulus-driven decision-making*

44

45 In the concurrent discrimination task, subjects select randomly on the first exposure of an
46
47 item, having no initial valid information that can be used to predict the choice outcome. With
48
49 repetition, information is continuously gathered regarding the correct and incorrect outcomes and
50
51 subsequently used to improve performance. After multiple rounds, this improved performance is
52
53 manifested as faster and more accurate selection of the rewarded item in each pair (Fig. 1b). In a
54
55 sense, decision-making becomes proceduralized.
56
57
58
59
60
61
62
63
64
65

1
2
3
4 The apparent increase in decision-making efficiency that arises from repeated encoding
5
6 could be seen as a result of the integrated functioning of the performance monitoring and
7
8 evidence storage systems discussed above, as illustrated in Figure 4. Early in the task, when
9
10 choice accuracy is low and response times are slow (Fig. 1b), adequate decision evidence might
11
12 not yet have been encoded to drive confident choice selection. This is reflected in the high
13
14 decision thresholds and low drift-rates for the earlier rounds of the task, signifying a need to
15
16 gather and evaluate better (or more) evidence (Fig. 1c). Correspondingly, RPEs are frequent
17
18 early on, since EVs are unreliable (i.e., low quality evidence). Control regions such as the dorsal
19
20 caudate, anterior insula, anterior cingulate, and pre-SMA might monitor these frequent error
21
22 signals (RPE) to prospectively determine the degree of cognitive oversight needed to ensure
23
24 optimal performance based on recent performance history.
25
26
27
28
29
30

31 As experience with the choices builds, accuracy begins to climb and RTs decrease (Fig.
32
33 1b), accompanied by a relaxation of decision thresholds and an increase in drift-rates (Fig. 1c).
34
35 Thus, EVs become more reliable, representing improvements to evidence quality. Consequently,
36
37 unexpected outcomes become less frequent. After sufficient repetition, bottom-up information
38
39 from putative evidence providers and accumulators in the putamen, temporoparietal regions, and
40
41 SMA becomes reliable enough to guide decisions under relatively lenient decision criteria. The
42
43 skill that potentially develops through repeated choice selection places an emphasis on stimulus-
44
45 driven information (e.g., outcome predictions), which builds as a result of earlier control (e.g.,
46
47 high decision thresholds). Learning through feedback might therefore enhance the quality of
48
49 contributing decision evidence, so that bottom-up signals become more salient and reliable, and
50
51 thus sufficient to enact a choice without burdensome goal, strategy, and error monitoring.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 The striatum seems to play a particularly important role in this process. These roles are
5
6 well characterized by reinforcement learning (Knutson et al., 2005), wherein the caudate and
7
8 putamen contribute separately to learning processes (Seger and Cincotta 2005) and in the
9
10 formation of value-driven stimulus-response associations (Yin et al., 2004, 2006; Tricomi et al.,
11
12 2009). In the present study, EV and RPE relate to different aspects of decision-making. EV
13
14 captures a history of the feedback experienced on prior trials. RPE captures momentary learning
15
16 signals that reflect the outcome of the current trial. We found that the putamen seems to capture
17
18 EV (i.e., predicted outcome value), whereas the caudate seems to capture positive RPE signals
19
20 (i.e., learning signals). Thus, information held by the putamen is drawn upon as decision
21
22 evidence (reflected in drift-rate), supporting the development of habit-like responses as
23
24 information becomes well-learned. These stimulus-response associations arise as a result of
25
26 computations in the caudate. For instance, microstimulation of the caudate at the time of reward
27
28 (i.e., post-response feedback) can enhance the rate of associative learning (Williams and
29
30 Eskandar, 2006). Moreover, the caudate head is sensitive to flexible, short-term changes in
31
32 learning that might reflect errors between predictions and outcomes (Kim and Hikosaka, 2013).
33
34 Our findings suggest that the caudate might not directly influence decisions (via drift-rate), but
35
36 rather seems to respond to or monitor post-decision events. Between the dorsomedial and
37
38 dorsolateral striatum, then, learning is a process of outcome signaling and subsequent
39
40 information updating (Miyachi et al., 2002). Information about errors in outcome predictions
41
42 might be processed in the caudate and might subsequently influence the function of the putamen
43
44 related to information updating or utilization of a history of outcomes to inform choice-selection.
45
46
47
48
49
50
51
52
53
54

55 It is worth noting that a key limitation to these interpretations is the exclusion of negative
56
57 RPE trials from the analysis. While these trials have clear theoretical value, they were
58
59
60
61
62
63
64
65

undersampled in this study's design. Thus, the present task cannot be leveraged with a regression analysis to produce reliable and confident results related to these negative RPE signals. A task that features frequent errors as a critical part of learning would be much better suited for investigating negative RPEs with confidence. Concurrent discrimination, however, wherein many subjects have perfect performance in later rounds, is more about repetition of positive choices.

4.4 Contributions of the medial temporal lobe

Given the nature of the task, it was expected that declarative memory regions such as the hippocampus and medial temporal lobe would map onto the EV and drift-rate intersection. While these regions were absent in this intersection, it should be noted that the parahippocampal gyrus was present in both the drift-rate and EV maps. It would make sense that declarative memory would play a role in concurrent discrimination learning. Such a system would be well suited to supply key contextual information about episodic associations between items in a present pair to support encoding for individual decision events (Burgess et al., 2002; Kirschoff et al., 2000). There are several possible explanations for the apparent absence of the influence of a declarative memory system on improvements in deterministic decision-making. First, given that the parahippocampal regions on each map were in different hemispheres (albeit at homologous locations), it is possible that the thresholds used to define the regions were too strict and washed out any overlapping effects (i.e., false negative). This is an unfortunate tradeoff for the selection of any threshold value. Second, it is possible that reinforcement learning algorithms cannot capture behavior related to hippocampal memory. This seems unlikely given that we did identify parahippocampal regions whose activity correlated with EV and that the learning rate of the model was high (i.e., single-trial learning was possible within the model). However, it is possible

1
2
3
4 that the model-free algorithm we chose for our analyses was not an adequate choice to fully
5
6 capture hippocampal influences and instead a model-based reinforcement learning approach
7
8 would be more appropriate (Doll et al., 2015; Johnson and Redish, 2005). A third possibility is
9
10 that the context of the task overburdens declarative memory and places emphasis on more
11
12 traditional reinforcement learning circuits (such as striatal procedural memory). For instance, our
13
14 task used 50 items, which is far more than the typical 8-12 items used in the work with human
15
16 amnesics (for example, Hood et al., 1999; Bayley et al., 2005). Memorizing associations for 8
17
18 items can be supported by declarative memory, but it is less clear how such a system might
19
20 behave in the context of 50 items. Moreover, feedback or reward context might play a role in the
21
22 engagement of the hippocampal memory system, since our task featured salient visual feedback
23
24 while other permutations of discrimination learning tasks have used physical monetary feedback,
25
26 such as dimes (Hood et al., 1999).
27
28
29
30
31
32

33 These issues of task context and of tradeoffs between declarative and procedural memory
34
35 have been highlighted in much of the literature on concurrent discrimination learning. For
36
37 instance, performance on concurrent discrimination tasks with a compromised declarative
38
39 memory system can sometimes be spared (Ridley et al., 1989; Gaffan et al., 1991). This might be
40
41 influenced by task design factors, such as including long retention intervals between decision
42
43 episodes (Malamut et al., 1984). Moreover, preserved performance ability seems to be mediated,
44
45 at least in part, by the striatum (Teng et al., 2000). However, when larger areas of the temporal
46
47 lobe are damaged (such as the parahippocampal gyrus), performance is almost entirely abolished,
48
49 suggesting that some aspects of declarative, associative, or recognition memory are necessary for
50
51 learning to occur (Buffalo et al., 1998; Buffalo et al., 1999; Chudasama et al., 2008). Similar
52
53 mixed results can be found in humans, wherein robust habit learning seems to support learning in
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 amnesics in some instances (Bayley et al., 2005) but not in others (Hood et al., 1999). Likewise,
5
6 patients with deficient procedural learning capacity due to Parkinson's disease, seem to require a
7
8 declarative awareness of cue-reward relationships in order to succeed on concurrent
9
10 discrimination learning (Moody et al., 2010). This underscores the importance of the basal
11
12 ganglia for this type of learning, but also highlights the idea that multiple systems can and may
13
14 interact. In the end, it seems that concurrent discrimination learning is a product of interactions
15
16 between both hippocampal and basal ganglia systems, but these interactions are modulated by
17
18 contextual and structural elements of the task (e.g., Turchi et al., 2010).
19
20
21
22
23
24
25

26 *4.5 Conclusions*

27
28 In an integrative approach to investigating the neural substrates of feedback and decision-
29
30 making, our findings outline the cognitive processes and neural correlates that support an
31
32 intersection of feedback-based learning and decision-making. Feedback carries critical outcome
33
34 information that influences subsequent decision-making in numerous ways (Fig. 4). Outcome
35
36 monitoring regions in the pre-SMA, anterior insula, anterior cingulate, dlPFC, and dorsal caudate
37
38 nucleus help to translate external information about whether behavior is meeting expectations
39
40 into appropriate adjustments to control over decision-making. At the same time, evidence-
41
42 providing regions in the putamen, temporoparietal regions, and SMA allow mnemonic
43
44 information gathered from past decisions to be applied to a current decision-making episode.
45
46 Taken together, these findings highlight the critical influence of feedback on performance
47
48 monitoring and the accrual of mnemonic information about choice value. In particular, the
49
50 findings outline an intersection of neural substrates that are important to both feedback-based
51
52 learning and decision-making. As this learning occurs, evidence quality improves, leading to a
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 shift in processing from controlled, top-down decision-making to proceduralized, stimulus-
5
6 driven choice selection, each mediated by separable neural substrates.
7
8
9

10 11 12 13 14 15 16 17 18 **6. Acknowledgements**

19 This work was supported by the National Science Foundation (Grant SBE-0354420 Supplement
20
21 0839229 to J.A.F.) and the National Institutes of Health (R21 MH106317). The authors thank
22
23
24 Corrie Durisko, Abraham Snyder, Mark McAvoy, Max Novelli, and Mark Vignone.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

7. References

- Badre, D., Lebrecht, S., Pagliaccio, D., Long, N.M., Scimeca, J.M., 2014. Ventral striatum and the evaluation of memory retrieval strategies. *Journal of Cognitive Neuroscience* 26 (9), 1928-1948.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R., 2007. The English Lexicon Project. *Behav Res Methods* 39, 445-459. doi:10.3758/BF03193014
- Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47, 129-141. doi:10.1016/j.neuron.2005.05.020
- Bayley, P.J., Frascino, J.C., Squire, L.R., 2005. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* 436 (28), 550-553. doi:10.1038/nature03857
- Bogacz, R., Wagenmakers, E-J., Forstmann, B.U., Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10-16. doi:10.1016/j.tins.2009.09.002
- Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402, 179-181. doi:10.1038/46035
- Bowman, N.E., Kording, K.P., Gottfried, J.A., 2012. Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron* 75, 916-927. doi:10.1016/j.neuron.2012.06.035

- 1
2
3
4 Buffalo, E.A., Stefanaci, L., Squire, L.R., Zola, S.M., 1998. A reexamination of the concurrent
5
6 discrimination learning task: the importance of anterior inferotemporal cortex, area TE.
7
8 Behavioral Neuroscience 112 (1), 3-14.
9
- 10
11 Buffalo, E.A., Ramus, S.J., Clark, R.E., Teng, E., Squire, L.R., Zola, S.M., 1999. Dissociation
12
13 between the effects of damage to perirhinal cortex and area TE. Learning and Memory 5,
14
15 572-599.
16
17
- 18
19 Burgess, N., Maguire, E.A., O'Keefe, J., 2002. The human hippocampus and spatial and episodic
20
21 memory. Neuron 35, 625-641. doi:10.1016/S0896-6273(02)00830-9
22
- 23
24 Burnham, K.P., Anderson, D.R., 2004. Multimodel inference – understanding AIC and BIC in
25
26 model selection. Sociological Methods & Research 33, 261-304.
27
28 doi:10.1177/0049124104268644
29
- 30
31 Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J.B., 2010. Frontal theta links prediction errors
32
33 to behavioral adaptation in reinforcement learning. NeuroImage 49, 3198-3209.
34
35 doi:10.1016/j.neuroimage.2009.11.080
36
37
- 38
39 Chudasama, Y., Wright, K.S., Murray, E.A. 2008. Hippocampal lesions in rhesus monkeys
40
41 disrupt emotional responses but not reinforcer devaluation effects. Biol Psychiatry 63,
42
43 1084-1091.
44
- 45
46 Coltheart, M., 1981. The MRC Psycholinguistic Database. Q J Exp Psychol-A 33, 497-505.
47
48 doi:10.1080/146407481084400805
49
- 50
51 Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp 8,
52
53 109-114.
54
- 55
56 Daw, N.D., Doya, K., 2006. The computational neurobiology of learning and reward. Curr Opin
57
58 Neurobiol 16, 199-204. doi:10.1016/j.conb.2006.03.006
59
60
61
62
63
64
65

- 1
2
3
4 Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences
5
6 on humans' choices and striatal prediction errors. *Neuron* 69, 1204-1215.
7
8
- 9 Doll, B.B., Jacobs, W.J., Sanfey, A.G., Frank, M.J., 2009. Instructional control of reinforcement
10
11 learning: a behavioral and neurocomputational investigation. *Brain Res* 1299, 74-94.
12
13 doi:10.1016/j.brainres.2009.07.007
14
15
- 16 Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D., Daw, N.D., 2015. Model-based choices
17
18 involve prospective neural activity. *Nature Neuroscience* 18 (5), 767-772.
19
20
- 21 Doya, K., 2000. Complementary roles of basal ganglia and cerebellum in learning and motor
22
23 control. *Curr Opin Neurobiol* 10, 732-739. doi:10.1016/S0959-4388(00)00153-7
24
25
- 26 Dunovan, K.E., Tremel, J.J., Wheeler, M.E., 2014. Prior probability and feature predictability
27
28 interactively bias perceptual decisions. *Neuropsychologia* 61, 210-221.
29
30 doi:10.1016/j.neuropsychologia.2014.06.024
31
32
- 33 Frank, M.J., Claus, E.D., 2006. Anatomy of a decision: striato-orbitofrontal interactions in
34
35 reinforcement learning, decision-making, and reversal. *Psychological Review* 113 (2), 300-
36
37 326.
38
39
- 40 Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T.V., Cavanagh, J.F., Badre, D., 2015.
41
42 fMRI and EEG predictors of dynamic decision parameters during human reinforcement
43
44 learning. *J Neurosci* 35, 485-494. doi:10.1523/jneurosci.2036-14.2015
45
46
47
- 48 Friston, K.J., Tononi, G., Reeke, G.N., Jr., Sporns, O., Edelman, G.M., 1994. Value-dependent
49
50 selection in the brain: simulation in a synthetic neural model. *Neuroscience* 59, 229-243.
51
52 doi:10.1016/0306-4522(94)90592-4
53
54
- 55 Gaffan, E.A., Gaffan, D., Hodges, J.R., 1991. Amnesia following damage to the left fornix and
56
57 to other sites: a comparative study. *Brain* 114 (3), 1297-1313.
58
59
60
61
62
63
64
65

- 1
2
3
4 Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable
5
6 neural prediction error signals underlying model-based and model-free reinforcement
7
8 learning. *Neuron* 66, 585-595. doi:10.1016/j.neuron.2010.04.016
9
10
11 Glimcher, P.W., 2003. The neurobiology of visual-saccadic decision making. *Annu Rev*
12
13 *Neurosci* 26, 133-179. doi:10.1146/annurev.neuro.26.010302.081134
14
15
16 Glimcher, P.W., 2011. Understanding dopamine and reinforcement learning: the dopamine
17
18 reward prediction error hypothesis. *PNAS USA* 108 (3), 15647-15654.
19
20
21 Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI.
22
23 *NeuroImage* 9 (4), 416-429.
24
25
26 Green, S.B., 1991. How many subjects does it take to do a regression analysis. *Multivariate*
27
28 *behavioral research* 26 (3), 499-510.
29
30
31 Hanes, D.P., Schall, J.D., 1996. Neural control of voluntary movement initiation. *Science* 274,
32
33 427-430. doi:10.1126/science.274.5286.427
34
35
36 Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W., Rangel, A., 2008. Dissociating the role of
37
38 the orbitofrontal cortex and the striatum in the computation of goal values and prediction
39
40 errors. *J Neurosci* 28 (22), 5623-5630.
41
42
43 Heekeren, H.R., Marrett, S., Bandettini, P.A., Ungerleider, L.G., 2004. A general mechanism for
44
45 perceptual decision-making in the human brain. *Nature* 431, 859-862.
46
47 doi:10.1038/nature02966
48
49
50 Hood, K.L., Postle, B.R., Corkin, S., 1999. An evaluation of the concurrent discrimination task
51
52 as a measure of habit learning: performance of amnesic subjects. *Neuropsychologia* 37,
53
54 1375-1386.
55
56
57
58
59
60
61
62
63
64
65

- Hunt, L.T., 2008. Distinctive roles for the ventral striatum and ventral prefrontal cortex during decision-making. *J Neurosci* 28 (35), 8658-8659.
- Jantzen, K.J., Steinberg, F.L., Kelso, J.A.S., 2002. Practice-dependent modulation of neural activity during human sensorimotor coordination: a functional magnetic resonance imaging study. *Neuroscience Letters* 332, 205-209.
- Jantzen, K.J., Steinberg, F.L., Kelso, J.A.S., 2004. Brain networks underlying human timing behavior are influenced by prior context. *PNAS USA* 101 (17), 6815-6820.
- Johnson, A., Redish, A.D., 2005. Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Networks* 18 (9), 1163-1171.
- Kayser, A.S., Buchsbaum, B.R., Erickson, D.T., D'Esposito, M., 2010. The functional anatomy of a perceptual decision in the human brain. *J Neurophysiol* 103, 1179-1194. doi:11.1152/jn.00364.2009
- Kayser, A.S., Erickson, D.T., Buchsbaum, B.R., D'Esposito, M., 2010. Neural representations of relevant and irrelevant features in perceptual decision making. *J Neurosci* 30, 15778-89. doi:10.1523/jneurosci.3163-10.2010
- Kim, H.F., Hikosaka, O., 2013. Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron* 79, 1-10. doi:10.1016/j.neuron.2013.06.044
- Kirchhoff, B.A., Wagner, A.D., Maril, A., Stern, C.E., 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J Neurosci* 20, 6173-6180.
- Knutson, B., Taylor, J., Kaufman, M., Peterson, R., Glover, G., 2005. Distributed neural representation of expected value. *J Neurosci* 25 (19), 4806-4812. doi:10.1523/jneurosci.0642-05.2005

- MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835-1838. doi:10.1126/science/288.5472.1835
- Malamut, B., Saunders, R., Mishkin, M., 1984. Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behavioral Neuroscience* 98 (5), 759-769.
- Miezin, F.M., Maccotta, L. Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage* 11, 735-759. doi:10.1006/nimg.2000.0568
- Moody, T.D., Chang, G.Y., Vanek, Z.F., Knowlton, B.J., 2010. Concurrent discrimination learning in Parkinson's Disease. *Behavioral Neuroscience* 124, 1-8. doi:10.1037/a0018414
- Miyachi, S., Hikosaka, O., Lu, X., 2002. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp Brain Res* 146, 122-126. doi:10.1007/s00221-002-1213-7
- Mulder, M.J., Wagenmakers E., Ratcliff, R., Boekel, W., Forstmann, B.U., 2012. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32 (7), 2335-2343. doi:10.1523/jneurosci.4156-11.2012
- O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329-337. doi:10.1016/S0896-6273(03)00169-7

- Ojemann, J.G., Akbudak, E., Snyder, A.Z. McKinstry, R.C., Raichle, M.E., Conturo, T.E., 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *NeuroImage* 6, 156-167. doi:10.1006/nimg.1997.0289
- Ollinger, J.M., Corbetta, M., Shulman, G.L., 2001. Separating processes within a trial in event-related functional MRI. *NeuroImage* 13, 218-229. doi:10.1006/nimg.2000.0710
- Oyama, K., Hernadi, I., Iijima, T., Tsutsui, K., 2010. Reward prediction error coding in dorsal striatal neurons. *J Neurosci* 30, 11447-11457. doi:10.1523/jneurosci.1719-10.2010
- Ploran, E.J., Nelson, S.M., Velanova, K., Donaldson, D.I., Petersen, S.E., Wheeler, M.E., 2007. Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *J Neurosci* 27, 11912-11924. doi:10.1523/jneurosci.3522-07.2007
- Ploran, E.J., Tremel J.J., Nelson, S.M., Wheeler, M.E., 2011. High quality but limited quantity perceptual evidence produces neural accumulation in frontal and parietal cortex. *Cereb Cortex* 21, 2650-2662. doi:10.1093/cercor/bhr055
- Poldrack, R.A., Prabhakaran, V., Seger, C.A., Gabrieli, J.D.E., 1999. Striatal activation during acquisition of a cognitive skill. *Neuropsychol* 13, 564-574. doi:10.1037/0894-4105.13.4.564
- Poldrack, R.A., Gabrieli, J.D.E., 2001. Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. *Brain* 124, 67-82. doi:10.1093/brain/124.1.67
- Poldrack, R.A., Sabb, F.W., Foerde, K., Tom, S.M. Asamow, R.F., Bookheimer, S.Y., Knowlton, B.J., 2005. The neural correlates of motor skill automaticity. *J Neurosci* 25, 5356-5364. doi:10.1523/jneurosci.3880-04.2005

- 1
2
3
4 Ratcliff, R., 1978. A theory of memory retrieval. *Psychol Review* 85, 59-108. doi:10.1037/0033-
5
6 295X.52.2.59
7
8
9 Ratcliff, R., Hasegawa, Y.T., Hasegawa, R.P., Smith, P.L., Segraves, M.A., 2007. Dual diffusion
10
11 model for single-cell recording data from the superior colliculus in a brightness-
12
13 discrimination task. *J Neurophysiol* 97, 1756-1774. doi:10.1152/jn.00393.2006
14
15
16 Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice
17
18 decision tasks. *Neural Comput* 20, 873-922. doi:10.1162/neco.2008.12-06-420
19
20
21 Ridley, R.M., Aitken, D.M., Baker, H.F., 1989. Learning about rules but not about reward is
22
23 impaired following lesions of the cholinergic projection to the hippocampus. *Brain*
24
25 *Research* 502 (2), 306-318.
26
27
28 Rieskamp, J., 2006. Perspectives of probabilistic inferences: reinforcement learning and an
29
30 adaptive network compared. *J Exp Psychol Learn Mem Cogn* 32, 1355-1370.
31
32 doi:10.1037/0278-7393.32.6.1355
33
34
35 Roitman, J.D., Shadlen, M.N., 2002. Response of neurons in the lateral intraparietal area during
36
37 a combined visual discrimination reaction time task. *J Neurosci* 22, 9475-9489.
38
39
40 Schultz, W., Dayan, P., Montague, R., 1997. A neural substrate of prediction and reward.
41
42 *Science* 275, 1593-1599. doi:10.1126/science.275.5306.1593
43
44
45 Schultz, W., Tremblay, L., Hollerman, J.R., 1998. Reward prediction in primate basal ganglia
46
47 and frontal cortex. *Neuropharmacology* 37, 421-429. doi:10.1016/S0028-3908(98)00071-9
48
49
50 Schultz, W., Tremblay, L., Hollerman, J.R., 2000. Reward processing in primate orbitofrontal
51
52 cortex and basal ganglia. *Cerebral Cortex* 10, 272-283. doi:10.1093/cercor/10.3.272
53
54
55 Seger, C.A., Cincotta, C.M., 2005. The roles of the caudate nucleus in human classification
56
57 learning. *J Neurosci* 25, 2941-2951. doi:10.1523/jneurosci.3401-04.2005
58
59
60
61
62
63
64
65

- 1
2
3
4 Seger, C.A., Peterson, E.J., Cincotta, C.M., Lopez-Paniagua, D., Anderson, C.W., 2010.
5
6 Dissociating the contributions of independent corticostriatal systems to visual
7
8 categorization learning through the use of reinforcement learning modeling and Granger
9
10 causality modeling. *NeuroImage* 50, 644-656. doi:10.1016/j.neuroimage.2009.11.083
11
12
13
14 Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal
15
16 cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86, 1916-1936.
17
18
19 Smith, A.B., Taylor, E., Brammer, M., Rubia, K., 2004. Neural correlates of switching set as
20
21 measured in fast, event-related functional magnetic resonance imaging. *Human Brain*
22
23 *Mapping* 21, 247-256.
24
25
26 Smith, D., Sip, K.E., Delgado, M.R., 2015. Functional connectivity with distinct neural networks
27
28 tracks fluctuations in gain/loss framing susceptibility. *Human Brain Mapping* 36, 2743-
29
30 2755.
31
32
33
34 Snyder, A.Z., 1996. Difference image versus ratio image error function forms in PET-PET
35
36 realignment. In: Bailey, D., Jones, T. (Ed.), *Quantification of brain function using PET*.
37
38 Academic, San Diego.
39
40
41 Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of
42
43 model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 64, 583-616.
44
45
46 Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2014. The deviance information
47
48 criterion: 12 years on. *J R Stat Soc Ser B Stat Methodol* 76, 485-493.
49
50 doi:10.1111/rssb.12062
51
52
53 Spoormaker, V.I., Andrade, K.C., Schröter, M.S., Sturm, A., Goya-Maldonado, R., Sämann,
54
55 P.G., Czisch, M., 2011. The neural correlates of negative prediction error signaling in
56
57 human fear conditioning. *NeuroImage* 54 (3), 2250-2256.
58
59
60
61
62
63
64
65

- 1
2
3
4 Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction. MIT Press, Cambridge,
5
6 MA.
7
8
9 Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional
10
11 proportional system: an approach to cerebral imaging. Georg Thieme, Stuttgart, New York.
12
13
14 Teng, E., Stefanacci, L., Squire, L.R., Zola, S.M., 2000. Contrasting effects on discrimination
15
16 learning after hippocampal lesions and conjoint hippocampal-caudate lesions in monkeys. J
17
18 Neurosci 20(10), 3853-3863.
19
20
21 Tremel, J.J., Wheeler, M.E., 2015. Content-specific evidence accumulation in inferior temporal
22
23 cortex during perceptual decision-making. NeuroImage 109, 35-49.
24
25 doi:10.1016/j.neuroimage.2014.12.072
26
27
28 Tricomi, E., Balleine, B.W., O'Doherty, J.P., 2009. A specific role for posterior dorsolateral
29
30 striatum in human habit learning. European Journal of Neuroscience 29 (11), 2225-2232.
31
32
33 Tricomi, E., Fiez, J.A., 2008. Feedback signals in the caudate reflect goal achievement on a
34
35 declarative memory task. NeuroImage 41, 1154-1167.
36
37 doi:10.1016/j.neuroimage.2008.02.066
38
39
40 Tricomi, E., Fiez, J.A., 2012. Information content and reward processing in the human striatum
41
42 during performance of a declarative memory task. Cogn Affect Behav Neurosci 12, 361-
43
44 372. doi:10.3758/s13415-011-0077-3
45
46
47
48 Turchi, J., Devan, B., Yin, P., Sigrist, E., Mishkin, M., 2010. Pharmacological evidence that both
49
50 cognitive memory and habit formation contribute to within-session learning of concurrent
51
52 visual discriminations. Neuropsychologia 48 (8), 2245-2250.
53
54
55
56
57
58
59
60
61
62
63
64
65

- Ullsperger, M., Harsay, H.A., Wessel, J.R., Ridderinkhof, K.R., 2010. Conscious perception of errors and its relation to the anterior insula. *Brain Struct Funct* 214, 629-643. doi:10.1007/s00429-010-0261-1
- Valenstein, E., Bowers, D., Verfaellie, M., Heilman, K.M., Day, A., Watson, R.T., 1987. Retrosplenial amnesia. *Brain* 110, 1631-1646. doi:10.1093/brain/110.6.1631
- Wagner, A.D., Shannon, B.J., Kahn, I., Buckner, R.L., 2005. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci* 9, 445-453. doi:10.1016/j.tics.2005.07.001
- Watkins. C.J.C.H., Dayan, P., 1992. Q-Learning. *Machine Learning* 8, 279-292.
- Wheeler, M.E., Buckner, R.L., 2004. Functional-anatomic correlates of remembering and knowing. *NeuroImage* 21, 1337-1349. doi:10.1016/j.neuroimage.2003.11.001
- Wheeler, M.E., Petersen, S.E., Nelson, S.M., Ploran, E.J., Velanova, K., 2008. Dissociating early and late error signals in perceptual recognition. *J Cogn Neurosci* 20, 2211-2225. doi:10.1162/jocn.2008.20155
- Wheeler, M.E., Woo, S.G., Ansel, T., Tremel, J.J., Collier, A.L., Velanova, K., Ploran, E.J., Yang, T. The strength of gradually accruing probabilistic evidence modulates brain activity during a categorical decision. *J Cogn Neurosci* 27, 705-719. doi:10.1162.jocn_a_00739
- Wiecki, T.V., Sofer, I., Frank, M.J., 2013. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in python. *Front Neuroinform* 7, 14. doi:10.3389/fninf.2013.00014.
- Williams, Z.M., Eskandar, E.N., 2006. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nature Neurosci* 9, 562-568. doi:10.1038/nn1662
- Woo, C.W., Koban, L., Kross, E., Linquist, M.A., Banich, M.T., Ruzic, L., Andrews-Hanna, J.R., Wager, T.D., 2014. Separate neural representations for physical pain and social rejection. *Nature communications* 5, 1-12.

- 1
2
3
4 Yang, T., Shadlen, M.N., 2007. Probabilistic reasoning by neurons. *Nature* 447, 1075-1080.
5
6 doi:10.1038/nature05852
7
8
9 Yin, H.H., Knowlton, B.J., Balleine, B.W., 2004. Lesions of the dorsolateral striatum preserve
10 outcome expectancy but disrupt habit formation in instrumental learning. *European Journal*
11 of Neuroscience 19 (1), 181-189.
12
13
14
15
16 Yin, H.H., Knowlton, B.J., Balleine, B.W., 2006. Inactivation of dorsolateral striatum enhances
17 sensitivity to changes in the action-outcome contingency in instrumental conditioning.
18
19 Behavioral Brain Research 166 (2), 189-196.
20
21
22
23
24 Zhang, J., Rowe, J.B., 2014. Dissociable mechanisms of speed-accuracy tradeoff during visual
25 perceptual learning are revealed by a hierarchical drift-diffusion model. *Front Neurosci* 8,
26
27 69. doi:10.3389/fnins.2014.00069
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Captions

Table 1. Estimated group-level parameters for drift diffusion models. a , decision threshold values (subscript indicates the Round number); v , mean-drift rate values; t_{er} , mean non-decision time; z , starting point; st , cross-trial variance in non-decision time; sv , cross-trial variance in drift-rate; sz , cross-trial variance in starting point; SE , standard error of the mean estimates across subjects.

Param	Group Mean	SE
a_1	1.347	0.018
a_2	1.924	0.020
a_3	1.807	0.020
a_4	1.745	0.021
a_5	1.683	0.021
a_6	1.664	0.022
a_7	1.652	0.023
a_8	1.526	0.021
v_1	-0.144	0.056
v_2	0.370	0.055
v_3	0.875	0.056
v_4	1.525	0.059
v_5	1.871	0.060
v_6	2.281	0.062
v_7	2.653	0.065
v_8	2.766	0.065
t_{er}	0.845	0.022
z	0.556	0.012
st	0.495	--
sv	0.556	--
sz	0.037	--

Table 2. Estimated model fits and learning rates for the reinforcement learning model. *Pseudo* R^2 , statistic illustrating degree of model fit (0.20 is generally a well-fit model); α , learning rate parameter.

Subject	Pseudo R^2	α
1	0.52	0.86
2	0.17	0.79
3	0.40	0.68
4	0.50	0.93
5	0.50	0.87
6	0.40	0.82
7	0.49	0.85
8	0.23	0.85
9	0.24	0.57
10	0.41	0.76
11	0.32	0.93
12	0.62	0.79
13	0.51	0.83
14	0.46	0.88
15	0.11	0.67
16	0.17	0.86

Table 3. Regions of interest whose activity correlated with modulations in Reward Prediction Errors. *x,y,z*, Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; *V_x*, region volume (in voxels); R, right; L, left; G, gyrus; S, sulcus; C, cortex; Ant, anterior; Inf, inferior; Med, medial; Mid, middle; Sup, superior; dlPFC, dorsolateral prefrontal cortex.

ROI	Anatomic Location	x	y	z	BA	V _x
1	Med Frontal G, pre-SMA	0	33	43	8	458
2	L Inf Occipital G	-24	-89	-12	19	237
3	L Inf Frontal G	-40	40	8	45	264
4	L Mid Frontal G	-32	53	16	46	202
5	L Med Frontal G, pre-SMA	-6	24	51	8	472
6	L Inf Frontal G	-48	4	25	44	427
7	R Ant Cingulate C	6	41	23	32	86
8	R Mid Frontal G, dlPFC	46	26	31	9	154
9	R Sup Frontal G	19	23	58	6	125
10	L Inf Frontal G	-49	14	27	44	233
11	L Caudate Head	-11	8	3	--	137
12	R Mid Frontal G	32	-3	45	6	83
13	L Ant Insula	-33	16	-1	13	260
14	R Inf Frontal G	48	8	32	44	149
15	R Fusiform G	27	-88	-9	19	189
16	R Fusiform G	45	-62	-11	19	122
17	L Inf Parietal Lobule	-45	-58	48	40	64
18	R Ant Insula	37	20	-1	13	177
19	R Mid Frontal G	34	54	12	10	216
20	L Cingulate G	-10	38	19	32	100
21	R Caudate Head	4	4	12	--	76
22	L Angular G	-40	-54	35	39	40

Table 4. Regions of interest whose activity correlated with modulations in the Expected Value covariate. *x,y,z*, Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; V_x, region volume (in voxels); R, right; L, left; G, gyrus; S, sulcus; C, cortex; Ant, anterior; Post, posterior; Inf, inferior; Med, medial; Mid, middle; Sup, superior; OFC, orbitofrontal cortex.

ROI	Anatomic Location	x	y	z	BA	V _x
1	L Mid Occipital G	-45	-66	10	19	119
2	R Mid Occipital G	15	-70	22	19	325
3	L Precuneus	-12	-48	51	7	228
4	R Parahippocampal G	15	-44	-4	30	207
5	R Cingulate G	1	-4	45	24	267
6	R Inf Parietal Lobule	59	-33	30	40	402
7	L Post Cingulate C	-2	-49	22	30	372
8	R Paracentral Lobule	1	-20	48	7	227
9	R Mid Insula	44	1	7	13	134
10	R Putamen	32	-16	15	--	127
11	R Mid Temporal G	52	-59	14	39	145
12	R Putamen	24	1	5	--	86
13	L Cuneus	-10	-87	23	19	129
14	R Precuneus	9	-42	45	7	64
15	R Med Frontal G, OFC	2	28	-12	11	56
16	R Thalamus	18	-20	6	--	59

Table 5. Regions of interest whose activity correlated with modulations in Decision Threshold.
x,y,z, Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; Vx, region
volume (in voxels); R, right; L, left; G, gyrus; S, sulcus; C, cortex; Ant, anterior; Post, posterior;
Inf, inferior; Med, medial; Mid, middle; Sup, superior; OFC, orbitofrontal cortex.

ROI	Anatomic Location	x	y	z	BA	Vx
1	R Ant Insula	33	19	-4	13	239
2	L Ant Insula	-29	23	2	13	188
3	R Precuneus	7	-71	36	7	337
4	R Med Frontal G, pre-SMA	2	27	44	8	701
5	L Fusiform G	-31	-85	-8	19	353
6	L Inf Parietal Lobule	-31	-58	43	40	535
7	R Sup Parietal Lobule	38	-57	50	7	149
8	R Mid Frontal G	41	12	41	9	290
9	L Precuneus	-17	-71	32	7	216
10	R Caudate Head	11	3	17	--	244
11	L Post Cingulate C	-1	-28	29	31	199
12	R Mid Occipital G	32	-79	-9	19	253
13	L Caudate Head	-8	2	12	--	138
14	R Angular G	30	-75	38	39	75
15	L Inf Parietal Lobule	-44	-37	45	40	90
16	L Mid Frontal G	-47	8	34	9	77
17	R Mid Occipital G	30	-94	7	19	54

Table 6. Regions of interest whose activity correlated with modulations in Drift-rate. x,y,z , Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; V_x , region volume (in voxels); R, right; L, left; G, gyrus; S, sulcus; C, cortex; Ant, anterior; Post, posterior; Inf, inferior; Med, medial; Mid, middle; Sup, superior; OFC, orbitofrontal cortex.

ROI	Anatomic Location	x	y	z	BA	V_x
1	R Inf Parietal Lobule	52	-29	27	40	570
2	L Inf Parietal Lobule	-56	-38	24	40	434
3	R Mid Temporal G	56	-58	9	39	88
4	R Mid Insula	38	-2	2	13	264
5	R Mid Temporal G	47	-38	-1	21	101
6	R Ant Cingulate C	4	-14	41	32	145
7	R Putamen	32	-16	9	--	249
8	L Post Insula	-38	-15	-1	13	190
9	L Parahippocampal G	-21	-39	-13	35	128
10	L Putamen	-28	-9	11	--	129
11	L Cingulate G	-7	1	45	24	174
12	L Precuneus	-4	-60	49	7	73

Table 7. Regions of interest defined by the intersection of Decision Threshold and Reward Prediction Error effects. *x,y,z*, Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; Vx, region volume (in voxels); R, right; L, left; G, gyrus; S, sulcus; Ant, anterior; Inf, inferior; Med, medial; Mid, middle.

ROI	Anatomic Location	x	y	z	BA	Vx
1	R Ant Insula	38	20	-3	13	61
2	L Ant Insula	-29	23	2	13	67
3	R Med Frontal G, pre-SMA	3	16	53	6	187
4	R Med Frontal G, pre-SMA	1	30	46	8	258
5	L Fusiform G	-29	-86	-11	19	100
6	R Caudate Head	8	7	12	--	24
7	R Ant Cingulate C	1	35	34	32	39
8	R Fusiform G	36	-76	-15	19	12
9	L Caudate Head	-8	2	12	--	18
10	R Inf Occipital G	24	-83	-5	19	25
11	L Mid Frontal G, dlPFC	-44	6	33	9	43

Table 8. Regions of interest defined by the intersection of Drift-rate and Expected Value effects. x,y,z , Talairach atlas coordinates of center of mass; BA, approximate Brodmann area; Vx, region volume (in voxels); R, right; L, left; G, gyrus; Inf, inferior; Mid, middle; Sup, superior; Ant, anterior.

ROI	Anatomic Location	x	y	z	BA	Vx
1	R Inf Parietal Lobule	56	-31	33	40	269
2	R Mid Temporal G	50	-57	12	39	22
3	R Mid Insula	37	-2	3	13	45
4	R Paracentral Lobule, SMA	4	-14	41	7	18
5	R Putamen	29	-15	9	--	47

Figure Captions

Figure 1. a, Concurrent discrimination task. In a set of 50 items, subjects are presented with a pair of words, one of which is designated the target (i.e., correct). Subjects choose the word and receive visual feedback. The set of 50 items is repeated 8 times total (8 Rounds), during which subjects learn the correct choices. b, Average accuracy across rounds (left panel) and response time across rounds (right panel) is plotted for the task participants (empirical) and for simulations of a drift-diffusion model (simulated) (see Table 1). Shaded area reflects standard error of the means. c, Threshold (left) and drift-rate (right) parameters across rounds of the drift-diffusion model fit from task behavior (see Table 1).

Figure 2. Activation map illustrating effects of reward prediction error (RPE) and expected value (EV) covariates derived from the reinforcement learning model. Z value indicates offset in mm from anterior commissure/posterior commissure atlas in Talairach space. Heat map scales for RPE (red to yellow) and EV (blue to green) reflect magnitude of the z-transformed t-statistic. Tables 3 and 4 list details of these illustrated regions.

Figure 3. Activation map illustrating effects of decision threshold and drift-rate covariates derived from the drift-diffusion model. Z value indicates offset in mm from ACPC axis in Talairach space. Heat map scales for decision threshold (red to yellow) and drift-rate (blue to green) reflect magnitude of the z-transformed t-statistic. Tables 5 and 6 list details of these illustrated regions.

Figure 4. a, Schematic illustrating the influences of feedback on a decision process. The lighter grey boxes indicate a theoretical process that is a specific focus of the present study, with a corresponding measure of interest indicated in parentheses (e.g., EV is a measure of decision evidence). The color-shaded areas represent a theoretical mapping of the two intersections of reinforcement learning and drift-diffusion parameters in the present study. EV, expected value; RPE, reward prediction error. b, Regions of interest corresponding to the Threshold + RPE (red) and Drift-rate + EV (blue) intersections. Z coordinates indicate the position of the axial slice in mm from ACPC line in Talairach space.

Figure 1

Figure 2

Figure 3

Figure 4