Behavioral/Cognitive

Frontal Theta Overrides Pavlovian Learning Biases

James F. Cavanagh,[1] **Ian Eisenberg,**[1] **Marc Guitart-Masip,**[2] **Quentin Huys,**[3,4] **and Michael J. Frank**[1,5,6]

[1]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02915, [2]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom, [3]Translational Neuroimaging Unit, Department of Biological Engineering, ETH and University of Zurich, 8092 Zurich, Switzerland, [4]Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital of Psychiatry Zurich, CH-8091 Zurich, Switzerland, and [5]Department of Psychiatry and [6]Brown Institute for Brain Science, Brown University, Providence, Rhode Island

Pavlovian biases influence learning and decision making by intricately coupling reward seeking with action invigoration and punishment avoidance with action suppression. This bias is not always adaptive—it can often interfere with instrumental requirements. The prefrontal cortex is thought to help resolve such conflict between motivational systems, but the nature of this control process remains unknown. EEG recordings of midfrontal theta band power are sensitive to conflict and predictive of adaptive control over behavior, but it is not clear whether this signal reflects control over conflict between motivational systems. Here we used a task that orthogonalized action requirements and outcome valence while recording concurrent EEG in human participants. By applying a computational model of task performance, we derived parameters reflective of the latent influence of Pavlovian bias and how it was modulated by midfrontal theta power during motivational conflict. Between subjects, those who performed better under Pavlovian conflict exhibited higher midfrontal theta power. Within subjects, trial-to-trial variance in theta power was predictive of ability to overcome the influence of the Pavlovian bias, and this effect was most pronounced in subjects with higher midfrontal theta to conflict. These findings demonstrate that midfrontal theta is not only a sensitive index of prefrontal control, but it can also reflect the application of top-down control over instrumental processes.

Introduction

Our prefrontal cortices allow us to deliberately overcome habitual or prepotent biases. However, some biases can exist even in novel or seemingly impartial situations. For example, innate Pavlovian biases facilitate reward-induced vigor and punishment-induced inhibition, yet these biases can actually hinder learning in instrumental conditions (Guitart-Masip et al., 2011, 2012b). Here we investigated whether prefrontal cortex can detect and resolve such conflict between separate motivational systems.

Motivated action selection is informed by at least three major processes that learn to associate stimuli with responses. Learning is primarily associated with two different instrumental processes: one reinforces rewarded actions and suppresses punished actions, resulting in stimulus-response behavior (Frank, 2005), and a second, more sophisticated prefrontal cortical operation involves an understanding of the consequences of actions and leads to goal-directed choices (Hampton et al., 2006). These systems are respectively referred to as "model-free" and "model-based", referring to the absence or presence of a "model" of the environ-

ment. Logically, organisms with these systems should be able to learn any type of stimulus-action pairing, yet in reality they can fail in reliable and predictable ways. This failure can be due to a third process: the influence of hard-coded Pavlovian responses to particular outcome expectations. Under Pavlovian influence, action selection and outcome valence are intricately interwoven.

Although in many circumstances Pavlovian influences may facilitate instrumental learning, they can also impair it by stifling the pairing of conflicting action and valence requirements (Dayan and Balleine, 2002). This was most clearly shown with Hershberger's (1986) chickens, which were unable to learn to move away from food to obtain it, and Holland's (1979) rats, which approached a light predictive of food even when that led to food omission. In these cases, Pavlovian biases did not simply interfere with correct performance, they actually led to the opposite behavior than the task required.

Nevertheless, unlike chickens, people can eventually learn conflicting instrumental contingencies. Whereas previous research has highlighted the neural mechanisms of Pavlovian influences on instrumental performance (Cardinal et al., 2002; Talmi et al., 2008; Prévost et al., 2012), little is known about the systems that overcome Pavlovian conflict. To investigate this topic, we used midfrontal EEG signals associated with conflict, learning, and control. Midfrontal theta-band power has been associated with the transient application of cognitive control to prevent impulsive responses (Cohen et al., 2009; Cavanagh et al., 2011). Therefore, frontal theta may reflect the recruitment of control, but only if a model-based system recognizes the conflict between motivational systems. However, other evidence indicates that

midfrontal theta covaries with reinforcement prediction error signals and can be predictive of subsequent learning and behavioral adjustments (Cavanagh et al., 2010, 2012a; van de Vijver et al., 2011). Therefore, this EEG signal appears to be influenced by both control (which could counter Pavlovian conflict) and value (which would covary with Pavlovian biases). By orthogonalizing action and valence, the current study was able to dissociate the influence of model-free (reflecting Pavlovian bias) and model-based (reflecting control over Pavlovian conflict) systems as indicated by frontal theta.

Materials and Methods

Participants. A total of 34 adults were recruited from the Brown University undergraduate subject pool and the Providence, Rhode Island, community to complete the experiment (20 males; median age, 24 years; SD, 4.44 years; range, 18–34 years). All participants had normal or corrected-to-normal vision, no history of neurological, psychiatric, or any other relevant medical problem, and were free from current psychoactive medication use. Participants were compensated $15 for completing the task, with a monetary bonus depending on total accuracy levels ($5 for achieving >50% total accuracy and $10 for achieving >66% accuracy).

Task: learning. The learning task was adapted from Guitart-Masip et al. (2011). Each trial consisted of three events: a cue, a target detection task, and an outcome (Fig. 1a). Before the experiment, participants were informed that each cue would either lead to reward or punishment based on their response and that no cue would lead to both. Participants were encouraged to explore both response options to best learn how to achieve the best outcome from each cue; they understood that they could either respond to the target ("Go") or they could withhold a response ("NoGo"). There were four different cues that predicted unique optimal combinations of action and outcome: Go-to-Win, Go-To-Avoid, NoGo-To-Win, and NoGo-to-Avoid, each of which was presented 40 times across two blocks with a break between.

Trials began with the display of a colored shape cue (1000 ms). Each cue was 70% predictive of the correct action to take (Go or NoGo) to gain the optimal outcome (reward or avoid punishment), whereas 30% or incorrect actions were reinforced. After a variable interval (250–2500 ms), the target detection stimulus appeared, which consisted of a white circle in the middle of the screen indicating that the subject could respond (Go) or not (NoGo). The circle disappeared if participants pressed the button or after 1000 ms. At 1000 ms after the offset of the circle, feedback was presented (2000 ms) indicating reward (green +$), punishment (red −$), or neutral (a yellow bar), which was used to indicate either no reward or punishment avoidance, depending on the condition. The intertrial interval consisted of a fixation cross for a variable interval (750–1500 ms).

In addition to these four probabilistic valenced conditions, there were two deterministic neutral conditions. Deterministic trials consisted of pictures of a hand indicating whether or not to press the response button. In these trials, participants were explicitly informed how to respond to the target detection task and were informed that there would be no outcome for these actions. These deterministic trials were intended to be used as motor action and inhibition contrasts, but were not used for the present analysis. Note that the task structure was shorter (40 vs 60 trials per condition) and harder (70% vs 80% reinforcing) than previous versions of this learning task (Guitart-Masip et al., 2012b).

Learning performance measures. Individual differences in performance styles were defined in relation to the hardest NoGo-to-Win condition in
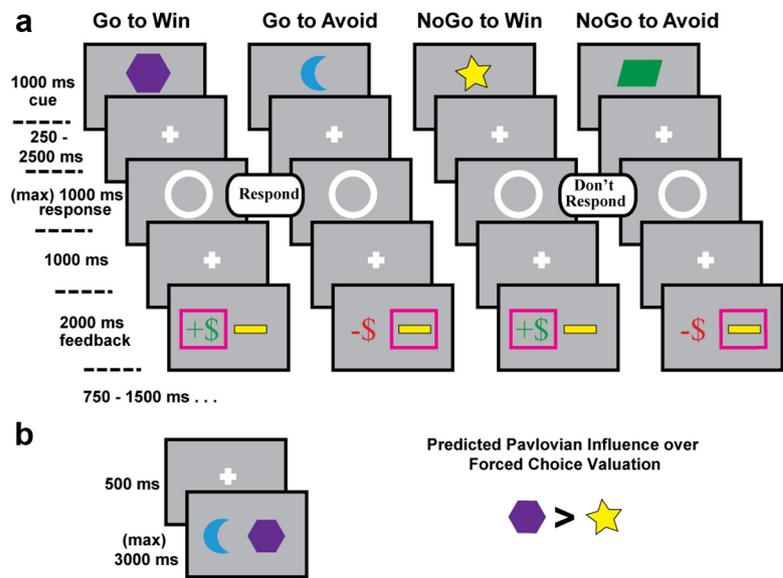

Figure 1. Tasks. *a*, Training phase. Four different cues were 70% predictive of unique optimal combinations of action and outcome: Go-to-Win, Go-To-Avoid, NoGo-To-Win, and NoGo-to-Avoid. In the context of a Pavlovian bias over performance, conditions with congruent action-outcome pairings should be easy to learn (Go-to-Win, NoGo-to-Avoid), whereas conditions with conflicting action-outcome pairings should be hard to learn (Go-to-Avoid, NoGo-to-Win). *b*, Posttraining transfer phase. Each cue was paired together in a two-alternative forced-choice testing phase; participants were told to select the most rewarding stimulus. It was hypothesized that participants would show a Pavlovian bias during this phase by selecting Go-to-Win more than NoGo-to-Win cues.

the second time block. If subjects successfully inhibited action on >65% of trials in this second block, they were considered "Learners" (*n* = 17); the rest of the subjects were labeled "Non-Learners" (*n* = 17). This categorical assignment facilitates an intuitive display of performance patterns, but we used continuous measures of learning for all important statistical analyses.

To condense the specific types of Pavlovian biases underlying performance differences across conditions, we devised the following reinforcement responsiveness metrics. To summarize across all blocks, measures of reward-based invigoration [(Go on Go-to-Win + NoGo-to-Win)/Total Go] and punishment-based suppression [(NoGo on Go-to-Avoid + NoGo-to-Avoid)/Total NoGo] were averaged into a single measure of Pavlovian Performance Bias (Figure 2b). Whereas approach and avoidance conditions are not motivationally identical, this measure effectively merges comparable instrumental outcome-action adaptations. Therefore, if a participant were to learn the conditions perfectly, his/her Pavlovian Performance Bias measure would be 50% because half of the conditions facilitated and the other half contradicted Pavlovian response styles. Higher scores on the measure of Pavlovian Performance Bias therefore reflect a greater dependence on Pavlovian biases during decision making.

Task: postlearning transfer phase. After the learning phase, participants also completed a novel forced-choice transfer phase after the task (this was not included in earlier studies using this task, but is similar to the transfer phase of Frank et al., 2004; Fig. 1b). Data from this transfer phase were not available for one participant (a Learner) who had to leave early. In this transfer phase, each of the predictive cues was paired with each of the others in a two-alternative forced choice scenario, and participants were told to select which cue was "more rewarding." No feedback was presented and each pairing was presented eight times. These choices were used to indicate relative preferences/valuations of the different cues to indicate their learned values outside of the instrumental learning environment. We reasoned that choice preferences in this phase would be indicative of Pavlovian biases in learned value, such that participants may assign a higher reward value for Go-to-Win than NoGo-to-Win cues. This hypothesized pattern of choices would reveal whether participants who do eventually learn the conflicting instrumental contingencies (i.e., Learners) nevertheless exhibit Pavlovian influences on value in their in-
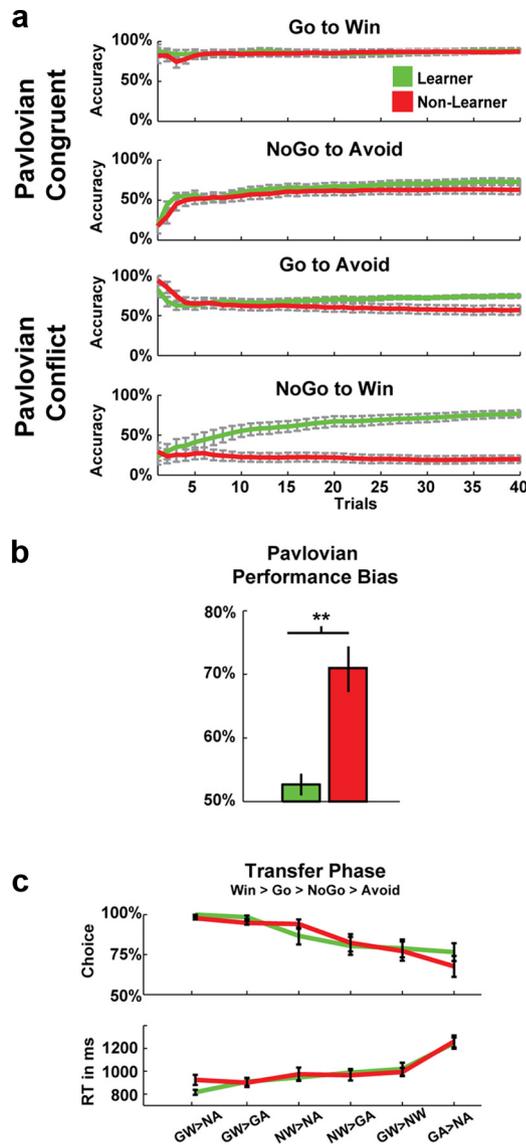
Figure 2. Task performance for Learners and Non-Learners as defined by accuracy in the latter half of the hardest NoGo-to-Win condition. ***a***, Participant performance during training was similar on easy, congruent conditions, but there were large differences in harder conditions characterized by Pavlovian conflict (Go-to-Avoid and NoGo-to-Win). ***b***, Measures of total invigoration on reward conditions and suppression on avoidance conditions were combined to create a single measure of Pavlovian Performance Bias, capturing an individual's aggregate tendency to commit an action in the presence of a reward-predictive cue and to withhold an action in the presence of a punishment-predictive cue across the entire experiment. ***c***, In a posttask forced choice transfer phase, subjects displayed a Win > Go > NoGo > Avoid ordering of preferences. This finding suggests that both groups had a Pavlovian influence during learning, but that the Learner group was somehow able to overcome this bias in the training phase. GW indicates Go-to-Win; GA, Go-to-Avoid; NW, NoGo-to-Win; and NA, NoGo-to-Avoid.

herent preferences. Such a Pavlovian influence over value-related forced choice may reveal whether the source of individual differences in Pavlovian biases resides in the mechanisms giving rise to the bias itself (Learners ≠ Non-Learners) or if subjects exhibit similar bias mechanisms but simply override them in the task conditions for which they are detrimental to performance (Learners = Non-Learners).

EEG recording and preprocessing. EEG was recorded using a 128-channel EGI system. EEG was recorded continuously with hardware filters set from 0.1 to 100 Hz, a sampling rate of 250 Hz, and an online vertex reference. Continuous EEG was epoched around the cues (−1500 to 5500 ms). Data were then visually inspected to identify bad channels to be interpolated and bad epochs to be rejected. Blinks were removed using

independent component analysis from EEGLab (Delorme and Makeig, 2004). The vertex site was reconstructed; data were then converted to current source density (Kayser and Tenke, 2006). Broadband ERPs were filtered from 0.5 to 20 Hz.

Time-frequency calculations were computed using custom-written MATLAB routines (Cavanagh et al., 2009). For condition-specific activities, time-frequency measures were computed by multiplying the fast-Fourier-transformed (FFT) power spectrum of single trial EEG data with the FFT power spectrum of a set of complex Morlet wavelets defined as a Gaussian-windowed complex sine wave: $e^{i2\pi tf}e^{-t^2/(2\times\sigma^2)}$, where t is time and f is the frequency (which increased from 1 to 50 Hz in 50 logarithmically spaced steps) that defines the width or "cycles" of each frequency band set according to $4/(2\pi f)$, and taking the inverse FFT. The end result of this process is identical to time-domain signal convolution provides estimates of instantaneous power (the magnitude of the analytic signal), defined as $Z[t]$ (power time series: $p(t) = \text{real}[z(t)]^2 + \text{imag}[z(t)]^2$). Each epoch was then cut in length (−500 to +1000 ms). Power was normalized by conversion to a decibel scale (10 * log10[power(t)/power(baseline)]), allowing a direct comparison of effects across frequency bands. For trial-to-trial analyses, EEG data were filtered from 4 to 8 Hz and Hilbert transformed to derive the single-trial theta power envelopes. The baseline for each frequency consisted of the average power from 300 to 200 ms before the onset of the cues.

Based on previous literature (Cavanagh et al., 2012b), the stimulus-locked theta band power burst over midfrontal sites (4–8 Hz, 175–350 ms) was a priori hypothesized to be the region of interest (ROI) involved in conflict and control. To verify this temporal, frequency, and spatial ROI using data-driven statistical tests, nonparametric Spearman's correlations of time-frequency space were used with behavioral or model-based parameters. To diminish the influence of outliers, trial-by-trial theta power values were sigmoid transformed before use in the computational model.

Computational modeling. An existing model of this task was used and refined to examine latent parameters thought to underlie individual differences in behavioral performance (Guitart-Masip et al., 2012b) and the degree to which these parameters were modified as a function of frontal theta. As in that study, models with increasing complexity (here labeled M1–M5) were assessed to determine whether they capture additional variance and provide better fits to the data (penalizing for additional complexity). Here we implement the novel advancement of investigating the influence of trial-by-trial theta power on action selection using three competing models (M6a, M6b, and M6c).

In all models, action values were estimated for each condition and a softmax choice function was used to predict the most likely action on each trial. The simplest model (M1) included two free parameters for scaling feedback sensitivity (ρ) and learning rate (ε). Reinforcements (r) took the form of (−1,0,1) depending on the condition, as follows:

$$r \in (-1,0,1)$$

State-action values (Q values) were updated according to the delta learning rule with feedback sensitivity (ρ) scaling the reinforcement value and the learning rate (ε) scaling the update term, as follows:

$$Q_t(a_t|s_t) = Q_{t-1}(a_t|s_t) + \varepsilon(\rho * r_t - Q_{t-1}(a_t|s_t))$$

Ensuing models included sequential additions, beginning with a third parameter in M2 to allow for irreducible noise (ξ) in action selection (to account for the possibility that some proportion of trials were not selected according to the model), as follows:

$$p(a|s) = \xi * \left[\frac{exp(Q(a_t|s_t))}{\sum exp(Q(q|s))} \right] + \frac{1-\xi}{2}$$

A fourth parameter was an overall bias to "Go" (b), regardless of valence, in M3, as follows:

$$Q_t(Go|s_t) = Q_t(Go|s_t) + b$$

A fifth parameter allowed for potentially different sensitivities to reward versus punishment (ρ_rew and ρ_pun) in M4. The critical sixth param-

eter in M5 specified the Pavlovian bias (π), which was the degree to which behavior is invigorated in response to stimuli that had positive learned value and is suppressed in response to stimuli that had negative learned value. In this model, the value V of each stimulus is learned as a function of reward history, then added to bias the action value $Q(\text{Go})$ in proportion to the Pavlovian bias as follows:

$$V_t(S_t) = V_{t-1}(S_t) + \varepsilon(\rho * r - V_{t-1}(S_t))$$

$$Q_t(\text{Go}|s_t) = Q_t(\text{Go}|s_t) + b + \pi * V_t(S_t)$$

Finally, to investigate whether midfrontal theta mitigated against Pavlovian bias, we investigated whether an Effect of Theta parameter (β) effectively weighted the trial-by-trial EEG theta power (θ_t) to alter the balance between the instrumental controller (Q) and the Pavlovian controller (V). Previous studies have indicated that the influence of midfrontal theta on cognitive control is primarily evident in conflict trials (Cavanagh et al., 2011), therefore this modulation was only modeled during conflict trials. We tested three models of this influence: M6a, M6b, and M6c.

M6a determined whether there was evidence for direct modulation of the Pavlovian influence (V) by theta power as follows:

$Q_t(\text{Go}|s_t)$

$$= \begin{cases} Q_t(\text{Go}|s_t) + b + (\pi + \beta * \theta_t) * V_t(S_t) & \text{if conflict} \\ Q_t(\text{Go}|s_t) + b + \pi * V_t(S_t) & \text{otherwise} \end{cases}$$

M6b determined whether there was evidence for a direct modulation of the instrumental contribution (Q) by theta power as follows:

$Q_t(\text{Go}|s_t)$

$$= \begin{cases} (1 - \beta * \theta_t) * Q_t(\text{Go}|s_t) + \pi * V_t(S_t) + b & \text{if conflict} \\ Q_t(\text{Go}|s_t) + b + \pi * V_t(S_t) & \text{otherwise} \end{cases}$$

$$Q_t(\text{NoGo}|s_t) = \begin{cases} (1 - \beta * \theta_t) * Q_t(\text{NoGo}|s_t) & \text{if conflict} \\ Q_t(\text{NoGo}|s_t) & \text{otherwise} \end{cases}$$

M6c determined whether there was evidence that theta power shifted control from the Pavlovian influence (V) toward an instrumental controller (Q) using an instrumental-Pavlovian trade-off parameter w instead of the Pavlovian influence π as follows:

$Q_t(\text{Go}|s_t)$

$$= \begin{cases} (1 - (w + \beta * \theta_t)) * Q_t(\text{Go}|s_t) \\ \quad + (w + \beta * \theta_t) * V_t(S_t) + b & \text{if conflict} \\ (1 - w) * Q_t(\text{Go}|s_t) + w * V_t(S_t) + b & \text{otherwise} \end{cases}$$

$Q_t(\text{NoGo}|s_t)$

$$= \begin{cases} (1 - (w + \beta * \theta_t)) * Q_t(\text{NoGo}|s_t) & \text{if conflict} \\ (1 - w) * Q_t(\text{NoGo}|s_t) & \text{otherwise} \end{cases}$$

In all of these models, if β was positive, this indicated that increasing theta power was associated with greater expression of Pavlovian biases, whereas if β was negative, it indicated that theta was a marker for the relative suppression of Pavlovian biases. In M6a, this is the consequence of $\beta * \theta_t$ directly modulating the Pavlovian influence. In M6b, reductions in the expression of Pavlovian contingencies are due to $\beta * \theta_t$ strengthening the instrumental component, whereas in M6c, they are the result of $\beta * \theta_t$ modulating the competition between the Pavlovian and the instrumental component.

As in previous publications of this model, an expectation-maximization procedure was used for hierarchical model estimation of group and individual subject parameters (Huys et al., 2011; Guitart-Masip et al., 2012b). Expectation-maximization recursively iterates model fitting to inform the group distribution for each model parameter, which is used as a prior for parameter maximization of each individual subject. Recursion finishes when consecutive iterations converge to near-identical parameter values. Model comparison used the integrated Bayesian Information Criterion (iBIC). Whereas the BIC provides an

estimate of the penalized *individual-level* likelihood of the data given a set of parameters, the iBIC estimates the penalized *group-level* likelihoods across the estimated *distribution* of the group-level hyperparameters. Lower iBIC values indicate a model that fits the data better, with a difference of 4–12 iBIC values suggesting positive evidence, 12–20 suggesting strong evidence, and above 20 suggesting very strong evidence (Kass and Raftery, 1995). As in previous uses of this model (Guitart-Masip et al., 2012b), feedback sensitivities and the Pavlovian bias were constrained to be positive and learning rates, the instrumental Pavlovian trade-off parameter w, and softmax noise were constrained to be between 0 and 1. All other parameters were unconstrained.

Results

Performance

Average performance accuracies followed qualitatively similar patterns as previous studies with this task (Guitart-Masip et al., 2012b), with good performance on Go-to-Win (accuracy mean = 0.88, SD = 0.13), somewhat equivalent performance on Go-to-Avoid (mean = 0.68, SD = 0.20), and NoGo-to-Avoid (mean = 0.68, SD = 0.21) and poorest performance on NoGo-to-Win (mean = 0.48, SD = 0.34). Figure 2a shows the group averages of the individual running accuracies in each condition. On Pavlovian congruent conditions, it is clear that all participants performed well. However, there was tremendous variance in performance on Pavlovian conflict conditions. Given that performance on Go-to-Avoid and NoGo-to-Win correlated with each other ($\rho_{(34)} = 0.43$, $p = 0.01$) and not with the congruent conditions (all $p > 0.16$), it is clear that Non-Learners were not simply idiosyncratically bad in some conditions, but rather subjects appeared to have reliable tendencies to rely on Pavlovian Biases. The summary measure of Pavlovian Performance Bias (Fig. 2b) differed between groups ($t_{(32)} = 4.62$, $p < 0.001$) and, as expected, were correlated with NoGo-to-Win accuracy ($\rho_{(34)} = -0.81$, $p < 0.001$) and Go-to-Avoid accuracy ($\rho_{(34)} = -0.67$, $p < 0.001$), effectively summarizing individual differences in the reliance on Pavlovian bias during the entire task.

Critically, both groups showed evidence for some Pavlovian influence over value learning in the posttask transfer phase (Fig. 2c). These findings reveal that all subjects displayed a clear Win > Go > NoGo > Avoid hierarchy of explicit preferences. Although the preference for Win > Avoid reflects the transfer phase instructions to select the "most rewarding" stimulus, a more subtle bias of Go > NoGo was also revealed in the pattern of choices. This finding is consistent with the idea that reward prediction errors invigorate action selection. There were no significantly different patterns between groups for any condition (all $t < 1.6$). Therefore, although Learners were able to successfully suppress Pavlovian biases during learning of the conflict conditions, they nevertheless exhibited the same choice preference for Go-to-Win over NoGo-to-Win and Go-to-Avoid over NoGo-to-Avoid as did the Non-Learners, despite the fact that these cues were similarly predictive of reward. These findings suggest that the difference between Learners and Non-Learners may not reside in the mechanisms giving rise to Pavlovian influences, but instead may reflect a differential ability to override such biases.

EEG

To investigate the influence of frontal theta power on Pavlovian conflict, we correlated the Pavlovian Performance Bias measure with the theta power difference between Pavlovian conflict and congruent conditions. This theta power contrast is orthogonal to action and reinforcement requirements (both conflict and congruent groupings involve one condition with a Go action and one with a NoGo action, and one condition with rewards and one
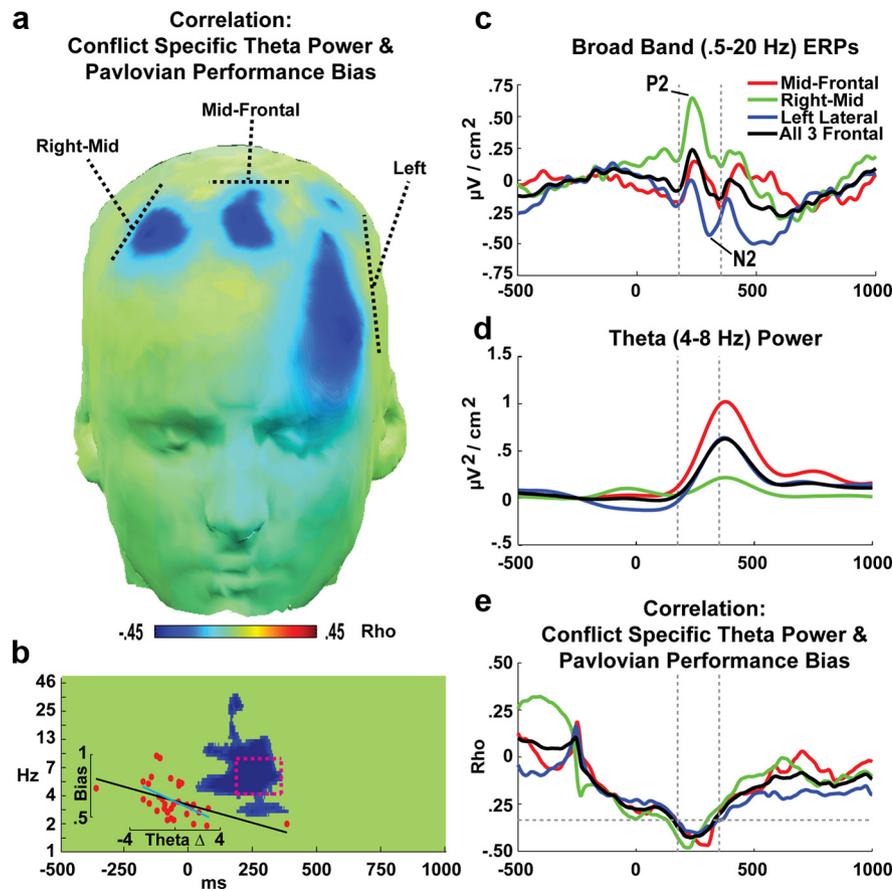
Figure 3. EEG activities in condition-specific effects and relationship with performance. **a**, Significant inverse correlations across a range of frontal sites between the cue-locked difference in theta (4 – 8 Hz) power for conflicting-congruent conditions and Pavlovian Performance Bias. This finding indicates that subjects with greater frontal theta to motivational conflict were characterized by a smaller Pavlovian bias and thus better performance in conflict trials. **b**, Pixel-wise correlations revealing that effects were prevalent around the boxed theta band ROI (175–350 ms, 4 – 8 Hz) over these frontal clusters. The inset scatterplot shows significant nonparametric correlations with (black fit line: $\rho_{(34)} = -0.55, p < 0.01$) or without (cyan fit line: $\rho_{(32)} = -0.50, p < 0.01$) outliers. **c**, ERPs demonstrate that the ROI time window occurs during the stimulus-locked P2–N2 complex. **d**, This time window specifically captured the waxing of the stimulus-locked theta band power burst. **e**, The correlation between conflict-congruent theta power and the Pavlovian Performance Bias was highly similar between sites, suggesting that these separate clusters reflect a common process. The horizontal dashed line represents the threshold for statistical significance. The discontinuity in the correlation at −250 ms is an artifact of baseline correction procedures.

with losses), providing a relatively clean measure of EEG activities associated with Pavlovian conflict rather than action or valence *per se*. There were no significant main effects of valence (Win or Avoid) or action (Go vs NoGo) in the ROI.

Figure 3a shows the topography of electrodes with a significant correlation between theta band power and Pavlovian Performance Bias scores. Major findings occurred in three broad electrode clusters over frontal cortex, referred to as midfrontal, right-mid, and left lateral. These regions are collapsed in Figure 3b to show the pixel-wise correlation of this performance measure with spectral differences. Significant effects were observed around the core temporal and frequency range of the ROI and these effects were replicated within each cluster independently. The inset shows the scatterplot of theta power with performance, revealing that the nonparametric correlations were significant with ($\rho_{(34)} = -0.55, p < 0.01$) or without ($\rho_{(32)} = -0.50, p < 0.01$) outliers. This ROI time range converges with the midfrontal P2-N2 complex of ERP components (Fig. 3c), which are characterized by a strong theta band spectral dynamic (Cavanagh et al., 2012b). Although average theta power differences varied across frontal clusters (Fig. 3d), the correlation values were similarly

strong within the ROI time windows (Fig. 3e). Interestingly, these correlations were maximal in the waxing of the theta power response.

In sum, convergent spatiotemporal frequency findings revealed that subjects with greater frontal theta power during early cue processing in response to Pavlovian conflict were less compromised by a Pavlovian Performance Bias. We next assessed whether trial-to-trial variations of theta *within* an individual were related to varying abilities to override Pavlovian biases. This question is most straightforwardly addressed in the context of the computational model fits to behavior by investigating whether trial-specific theta power from the same ROI influenced the expression of Pavlovian bias, the recruitment of instrumental contingencies, or both.

Computational modeling

Table 1 reveals that the stepwise addition of parameters in each model M1–M5 yielded increasingly better fits, as measured by iBIC, and how the novel model M6a provided the strongest improvement upon the fit of the data from next most complex model with only a Pavlovian Bias parameter (M5). M6a had highly similar parameters to M5; only the addition of the Effect of Theta parameter provided a better fit to the data. Figure 4a, b shows that within this best model (M6a), the Pavlovian Performance Bias (used in Fig. 2 and Fig. 3) was correlated with both the Pavlovian Bias parameter ($\rho_{(34)} = 0.60, p < 0.01$) and the Effect of Theta parameter ($\rho_{(34)} = 0.37, p < 0.05$). Moreover, the Pavlovian Bias parameter was highly similar between M5 and M6 ($\rho = .71, p <$

0.01), and was uncorrelated with the Effect of Theta parameter ($p = 0.85$), highlighting the fact that the Effect of Theta parameter accounted for unique variance in the improved model fit in M6.

The mean value for the Effect of Theta parameter (β) in M6a was significantly negative across the entire group (mean = −0.67, $t_{(33)} = 5.82, p < 0.01$), implying that trial-to-trial variations in theta negatively influenced within-subject Pavlovian biases. The Effect of Theta parameter was more negative in Learners (mean = −0.94) than in Non-Learners (mean = −0.40); these groups were significantly different from each other ($t_{(32)} = 2.55, p < 0.05$). Figure 4c reveals that interindividual differences in theta power increases in response to Pavlovian conflict (across participants) correlated with intraindividual abilities to use theta (across trials) to overcome Pavlovian biases. Table 1 also shows that alternative formulations in which theta promoted instrumental contingencies or explicitly modulated the trade-off between instrumental and Pavlovian influences provided inferior accounts of the data.

Therefore, subjects who were more likely to detect the presence of Pavlovian conflict exhibited higher midfrontal theta

Table 1. iBIC and parameter means (SD) for each model (M1–M5) and novel extensions (M6a–M6c) that examined the influence of trial-to-trial EEG theta power

	M1	M2	M3	M4	M5	M6a	M6b	M6c
iBIC	5613	5615	5379	5141	4926	4857	4947	4893
Feedback sensitivity (ρ)	4.05 (2.49)	4.67 (3.25)	4.95 (3.04)					
Reward sensitivity (ρ_rew)				12.82 (22.27)	6.78 (4.67)	6.86 (4.35)	7.85 (6.40)	9.81 (8.84)
Punishment sensitivity (ρ_pun)				3.76 (2.83)	4.81 (3.76)	6.26 (5.67)	4.77 (3.62)	9.36 (10.23)
Learning rate (ε)	0.28 (0.18)	0.29 (0.19)	0.29 (0.19)	0.29 (0.17)	0.28 (0.15)	0.23 (0.15)	0.32 (0.13)	0.27 (0.17)
Irreducible noise (ξ)		0.97 (0.02)	0.94 (0.08)	0.97 (0.02)	0.97 (0.01)	0.96 (0.03)	0.99 (0.01)	0.96 (0.03)
Go bias (b)			0.38 (0.62)	0.13 (0.85)	0.50 (0.60)	0.58 (0.61)	0.49 (0.59)	0.58 (0.63)
Pavlovian bias (π)					0.48 (0.77)	0.77 (0.75)	0.34 (0.64)	
Effect of theta (β)						−0.67 (0.67)	0.48 (0.73)	−0.32 (0.58)
Trade-off parameter (w)								0.31 (0.14)

power when conflict was high and were thus better at overcoming Pavlovian biases in those trials for which theta was particularly evident. The strong negative coefficient implies that on trials in which theta power was high, there was a diminished Pavlovian bias. This result converges with the findings from the behavioral transfer phase, which implied that even Learners exhibit a Pavlovian bias in their valuations, but are simply able to suppress that bias when they detect that it conflicts with the instrumental requirements of the task.

Discussion

This investigation revealed that conflict-induced midfrontal theta power is indicative of the ability to overcome Pavlovian biases when they conflict with instrumental requirements. This effect was observed both interindividually and intraindividually, where greater theta to conflict was associated with increased top-down adaptive control. These results are consistent with prior studies showing that bilateral inferior frontal gyri are involved when overcoming Pavlovian bias (Guitart-Masip et al., 2012b) and that conflict-related midfrontal theta influences the ability to prevent impulsive responding (Cavanagh et al., 2011).

Nature of Pavlovian biases
The results reported here replicate findings of a pervasive Pavlovian influence over instrumental performance (Talmi et al., 2008; Guitart-Masip et al., 2011, 2012a), including the finding that subjects vary widely in the expression of this influence (Guitart-Masip et al., 2012b). However, it remains unknown whether some subjects simply have a diminished Pavlovian influence over behavior or if these subjects actively overcome this bias through effort-based and goal-directed cognitive control mechanisms. Convergent evidence suggests the latter case.

All subjects demonstrated evidence for a learned coupling between action and valence (e.g., they preferred Go-to-Win over NoGo-to-Win) in the posttask transfer phase. Therefore, it appears that although Learners were able to successfully suppress Pavlovian biases during learning of the conflict conditions, they nevertheless exhibited the same forced choice preference as did the Non-Learners. This finding is particu-
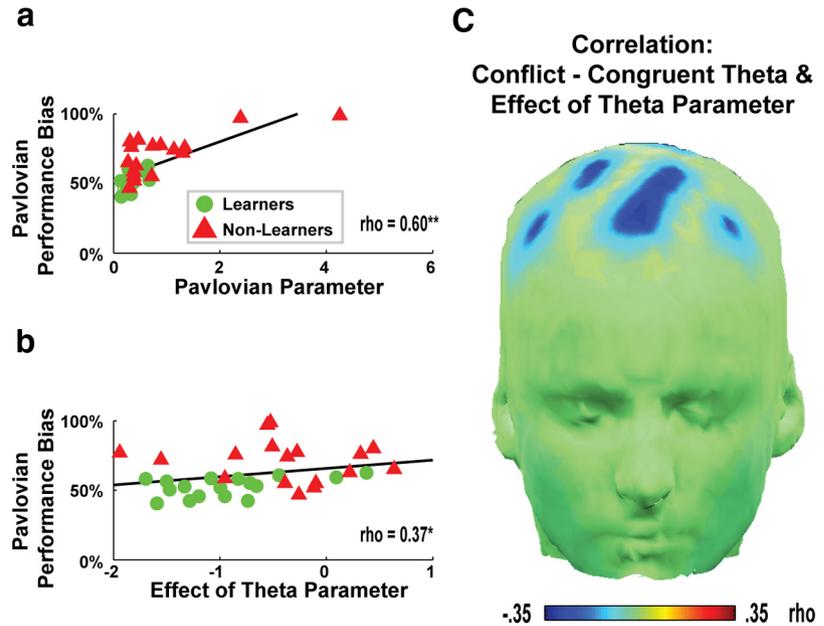

Figure 4. Parameters from the best fitting computational model (M6a). *a*, The Pavlovian Bias parameter estimated from the model significantly correlated with the Pavlovian Performance Bias metric used in Figure 2 and Figure 3. *b*, The parameter scaling the Effect of Theta in the model also significantly correlated with the Pavlovian Performance Bias metric, indicating that trial-to-trial variations in theta power diminished Pavlovian biases. *c*, The Effect of Theta parameter had a more adaptive influence (i.e., was more negative) in those subjects who had larger conflict-induced theta power.

larly notable because Learners had more experience with positive prediction errors in NoGo-to-Win than did Non-Learners, yet still showed the same action-biased preferences. Furthermore, within any individual Learner, trials with lower midfrontal theta responses to Pavlovian conflict during learning were associated with a greater propensity for Pavlovian biases on behavior. These findings suggest that Learners do not differ from Non-Learners in terms of the mechanisms giving rise to such biases (putatively related to *model-free* corticostriatal function), but rather in their ability to detect when these biases conflict with the rules of the task and need to be suppressed (putatively by *model-based* top-down prefrontal control; see also Guitart-Masip et al., 2012b).

Role of theta
The midfrontal effect occurred during the P2–N2 time range of the ERP, a temporally specific window known to be affected by conflict-induced cognitive control and expectation-induced mismatch, and known to have a strong spectral signature in the theta band (Hanslmayr et al., 2008; Cavanagh et al., 2012b) and a presumed generator in midcingulate cortex (Van Veen and Carter, 2002; Yeung et al., 2004; Hanslmayr et al., 2008). This is the same

approximate time period in the ERP that Holroyd et al. (2011) recently suggested was specifically sensitive to reward prediction errors, yet the current findings suggest that this previous effect may reflect general salience instead. Indeed, in this and other studies, midfrontal theta power appears to reflect a generic signal of the need for top-down control, not an axiomatic reward prediction error (Oliveira et al., 2007; Cavanagh et al., 2012a). However, very recent evidence has suggested that non-phase-locked power (as used here) may preferentially reflect violations of probability, whereas phase-locked ERP amplitude (as in Holroyd et al., 2011) may be primarily sensitive to valence (Hajihosseini and Holroyd, 2013). Clearly, more explicit hypothesis testing is needed to define the information content reflected within the ERP and constituent frequency bands and how this information may be differentially reflected in power versus phase activities.

It has been shown previously that midfrontal theta predicts behavioral slowing (Cavanagh et al., 2010) and switching (Cohen and Ranganath, 2007; van de Vijver et al., 2011) after prediction error; however, it was not known if this effect relied the operation of a model-free controller for generic slowing/switching or a model-based controller for adaptive behavioral adjustment. The findings from the present study suggest the latter case, given that conflict-induced mid frontal theta was implicated in the ability to overcome Pavlovian biases through both invigoration and inhibition of action. Comparison of the three novel models clearly favored an account in which trial-to-trial theta power suppressed Pavlovian influence (M6a) rather than promoting instrumental contingencies (M6b) or balancing the relative activity between the two (M6c). Based on theoretical and empirical findings, one candidate mechanism for this effect could be communication from the medial frontal cortex to the subthalamic nucleus via the hyperdirect pathway, which could raise the decision threshold to temporarily prevent the influence of striatal valuation signals on behavior (Frank, 2006; Cavanagh et al., 2011; Ratcliff and Frank, 2012; Zaghloul et al., 2012).

Remaining questions on the interplay between motivational systems

Pavlovian influence was modeled here and elsewhere as a bias in action invigoration/inhibition during response selection (Huys et al., 2011; Guitart-Masip et al., 2012b), but it remains possible that this effect also reflects biased learning. The architecture of the basal ganglia is structured to facilitate both learning and action selection in a manner biased by Pavlovian influences. Dopamine bursts to positive prediction errors facilitate plasticity along the D1-receptor-mediated direct pathway while suppressing the D2-receptor mediated indirect pathway, whereas the opposite is true for negative prediction errors (Frank, 2005; Gerfen and Surmeier, 2011; Kravitz et al., 2012). Therefore, actions are more likely to be invigorated and reinforced after positive prediction errors and more likely to be suppressed after negative prediction errors. As such, action biases could be accounted for by aberrant learning of action values, by a motivational alteration at the time of choice, or by an interaction of the two (Beeler et al., 2012). Given that these influences over motivation and learning are latent and possibly overlapping, it is difficult to parse the true nature of Pavlovian biases during action learning.

Guitart-Masip et al. recently demonstrated that increased activity in the striatum and substantia nigra/ventral tegmental area was associated with action invigoration and (2011, 2012a) and trial-by-trial action values (2012b). When the effects of valence on action learning were tested, there were no neural correlates of state predictive value or a Pavlovian interaction that could account for the observed Pavlovian biases during learning. However, this is possibly due to a high correlation between action and state values in this task, making it difficult to tease apart with fMRI. In addition, it is unknown whether biases in the posttask transfer phase of the present study reflect motivational, learning, or other influences that boost the apparent value of action over omission.

Although much of the evidence thus far favors a motivational account, it is likely that a more specific task and computational model will be required to test separable Pavlovian influences over motivation versus learning. For example, although the posttraining transfer phase data described here are generally supportive of a coupling between action and valence in value learning, the present model may not account for the full spectrum of choices. Specifically, participants reliably preferred Go-to-Avoid over NoGo-to-Avoid, which could not be explained by greater valuation *per se*, but may require models that impose a greater degree of learning from positive prediction errors after Go actions than after NoGo actions. Regardless of the specific mechanism of action/valence coupling, a similar top-down signal may be required to diminish it, suggesting that findings of increased frontal theta and bilateral inferior frontal gyrus (Guitart-Masip et al., 2012b) remain effective descriptions of the nature of model-based instrumental control.

Conclusion

In the present study, individual performances were characterized by a varied mixture of Pavlovian bias, instrumental learning, and model-based control. Through an innovative mixture of cognitive neuroscience and computational modeling we were able to determine the degree of Pavlovian bias over instrumental learning, and the nature of prefrontal control that was applied to ameliorate this bias. Our results suggest that midfrontal theta is a sensitive index of model-based prefrontal control over behavior.

References

Beeler JA, Frank MJ, McDaid J, Alexander E, Turkson S, Sol Bernandez MS, McGehee DS, Zhuang X (2012) A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. Cell Rep 2:1747–1761. CrossRef Medline

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352. CrossRef Medline

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

Cavanagh JF, Frank MJ, Klein TJ, Allen JJ (2010) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. Neuroimage 49:3198–3209. CrossRef Medline

Cavanagh JF, Wiecki TV, Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ (2011) Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat Neurosci 14:1462–1467. CrossRef Medline

Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ (2012a) Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. Cereb Cortex 22:2575–2586. CrossRef Medline

Cavanagh JF, Zambrano-Vazquez L, Allen JJ (2012b) Theta lingua franca: a common mid-frontal substrate for action monitoring processes. Psychophysiology 49:220–238. CrossRef Medline

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

Cohen MX, van Gaal S, Ridderinkhof KR, Lamme VA (2009) Unconscious errors enhance prefrontal-occipital oscillatory synchrony. Front Hum Neurosci 3:54. CrossRef Medline

Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. Neuron 36:285–298. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis

of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21 CrossRef Medline

Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. J Cogn Neurosci 17:51–72. CrossRef Medline

Frank MJ (2006) Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw 19:1120–1136. CrossRef Medline

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943. CrossRef Medline

Gerfen CR, Surmeier DJ (2011) Modulation of striatal projection systems by dopamine. Annual review of neuroscience 34:441–466. CrossRef Medline

Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJ, Dayan P, Dolan RJ, Duzel E (2011) Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J Neurosci 31:7867–7875. CrossRef Medline

Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ (2012a) Action controls dopaminergic enhancement of reward representations. Proc Natl Acad Sci U S A 109:7511–7516. CrossRef Medline

Guitart-Masip M, Huys QJ, Fuentemilla L, Dayan P, Duzel E, Dolan RJ (2012b) Go and no-go learning in reward and punishment: interactions between affect and effect. Neuroimage 62:154–166. CrossRef Medline

Hajihosseini A, Holroyd CB (2013) Frontal midline theta and N200 amplitude reflect complementary information about expectancy and outcome evaluation. Psychophysiology. Advance online publication. doi:10.1111/psyp.12040. Retrieved: Feb. 27, 2013. CrossRef Medline

Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360–8367. CrossRef Medline

Hanslmayr S, Pastötter B, Bäuml KH, Gruber S, Wimber M, Klimesch W (2008) The electrophysiological dynamics of interference during the Stroop task. J Cogn Neurosci 20:215–225. CrossRef Medline

Hershberger WA (1986) An approach through the looking-glass. Animal Learning and Behavior 14:443–451. CrossRef

Holland PC (1979) Differential effects of omission contingencies on various components of Pavlovian appetitive conditioned responding in rats. J Exp Psychol Anim Behav Process 5:178–193. CrossRef Medline

Holroyd CB, Krigolson OE, Lee S (2011) Reward positivity elicited by predictive cues. Neuroreport 22:249–252. CrossRef Medline

Huys QJ, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. PLoS Comput Biol 7:e1002028. CrossRef Medline

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795. CrossRef

Kayser J, Tenke CE (2006) Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin Neurophysiol 117:348–368. CrossRef Medline

Kravitz AV, Tye LD, Kreitzer AC (2012) Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci 15:816–818. CrossRef Medline

Oliveira FT, McDonald JJ, Goodman D (2007) Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J Cogn Neurosci 19:1994–2004. CrossRef Medline

Prévost C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural correlates of specific and general Pavlovian-to-instrumental transfer within human amygdalar subregions: a high-resolution fMRI study. J Neurosci 32:8383–8390. CrossRef Medline

Ratcliff R, Frank MJ (2012) Reinforcement-based decision making in corti-costriatal circuits: mutual constraints by neurocomputational and diffusion models. Neural Comput 24:1186–1229. CrossRef Medline

Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian-instrumental transfer. J Neurosci 28:360–368. CrossRef Medline

van de Vijver I, Ridderinkhof KR, Cohen MX (2011) Frontal oscillatory dynamics predict feedback learning and action adjustment. J Cogn Neurosci 23:4106–4121. CrossRef Medline

Van Veen V, Carter CS (2002) The timing of action-monitoring processes in the anterior cingulate cortex. J Cogn Neurosci 14:593–602. CrossRef Medline

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 111:931–959. CrossRef Medline

Zaghloul KA, Weidemann CT, Lega BC, Jaggi JL, Baltuch GH, Kahana MJ (2012) Neuronal activity in the human subthalamic nucleus encodes decision conflict during action selection. J Neurosci 32:2453–2460. CrossRef Medline