The Feedback-related Negativity Codes Components of Abstract Inference during Reward-based Decision-making

Andrea M. F. Reiter[1,2,3], Stefan P. Koch[4], Erich Schröger[2], Hermann Hinrichs[5,6], Hans-Jochen Heinze[5,6], Lorenz Deserno[1,4,6], and Florian Schlagenhauf[1,4]

Abstract

■ Behavioral control is influenced not only by learning from the choices made and the rewards obtained but also by "what might have happened," that is, inference about unchosen options and their fictive outcomes. Substantial progress has been made in understanding the neural signatures of direct learning from choices that are actually made and their associated rewards via reward prediction errors (RPEs). However, electrophysiological correlates of abstract inference in decision-making are less clear. One seminal theory suggests that the so-called feedback-related negativity (FRN), an ERP peaking 200–300 msec after a feedback stimulus at frontocentral sites of the scalp, codes RPEs. Hitherto, the FRN has been predominantly related to a so-called "model-free" RPE: The difference between the observed outcome and what had been expected. Here, by means of computational modeling of choice behavior, we show that individuals employ abstract, "double-update" inference on the task structure by concurrently tracking values of chosen stimuli (associated with observed outcomes) and unchosen stimuli (linked to fictive outcomes). In a parametric analysis, model-free RPEs as well as their modification because of abstract inference were regressed against single-trial FRN amplitudes. We demonstrate that components related to abstract inference uniquely explain variance in the FRN beyond model-free RPEs. These findings advance our understanding of the FRN and its role in behavioral adaptation. This might further the investigation of disturbed abstract inference, as proposed, for example, for psychiatric disorders, and its underlying neural correlates. ■

INTRODUCTION

A core function of human decision-making is to evaluate the motivational significance of ongoing events to weigh different decision options and to guide future decisions accordingly. There is growing consensus that individuals make decisions by computing decision values of potential options, which are then compared to make a choice (Sokol-Hessner, Hutcherson, Hare, & Rangel, 2012; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011; Rangel & Hare, 2010). It is thereby indispensable for an agent to keep track of choice values of options that were actually taken. This can be achieved via updating chosen values by reward prediction errors (RPEs), which result from comparing the observed outcome with what has been expected. However, the idea that an agent also learns from "what might have happened," that is, abstract inference regarding unchosen options and their fictive outcomes, has recently sparked considerable interest (Fischer & Ullsperger, 2013; Boorman, Behrens, & Rushworth, 2011; Boorman, Behrens, Woolrich, & Rushworth, 2009). Concurrent coding of multiple actions and their outcomes is suggested to enhance the efficiency of reinforcement learning (Takahashi et al., 2013; Abe & Lee, 2011; Boorman et al., 2011), and disturbed abstract inference on "what might have happened" has been implicated in psychopathological states, for example, in animal models of addiction (Lucantonio, Takahashi, et al., 2014; Lucantonio, Stalnaker, Shaham, Niv, & Schoenbaum, 2012).

In this study, we employ a probabilistic decision-making task and focus on simultaneous updating of chosen and unchosen decision values. In this task, decision values were anticorrelated such that the drop in one value directly implied the rise of the other (e.g., compare Schlagenhauf et al., 2013; Gläscher, Hampton, & O'Doherty, 2009; Hampton, Bossaerts, & O'Doherty, 2006). Thus, decision values for the options at hand can be inferred not only by action–reward pairings but also by abstract inference based on the anticorrelated task structure (Bromberg-Martin, Matsumoto, Hong, & Hikosaka, 2011). The standard reinforcement learning account only updates the value of a chosen stimulus via an RPE. This is referred to as a "model-free" RPE because it neglects the structure of the environment. Here, the anticorrelation of the two values. Although no outcome is delivered for the unchosen

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [2]University of Leipzig, [3]Technische Universität Dresden, [4]Charité-Universitätsmedizin Berlin, [5]Leibniz Institute for Neurobiology, Magdeburg, Germany, [6]Otto-von-Guericke University Magdeburg

option, via abstract inference on the reward statistics of the experimental environment, the agent could make use of the task structure and concurrently "double-update" the values of chosen and unchosen options. The model-free RPE is thereby modified by an abstract inference component. Double-updating optimizes learning even in simple but dynamic binary choice environments (Dolan & Dayan, 2013; Schlagenhauf et al., 2013; Li & Daw, 2011; Hampton et al., 2006).

In the current study, we aimed to define electrophysiological signatures of incorporating such abstract counterfactual inference on the task structure into the decision-making process. In a pioneering theory, Holroyd and Coles (2002) proposed that model-free RPEs linearly scale with the so-called feedback-related negativity (FRN). The FRN is an ERP peaking at frontocentral electrodes about 200–300 msec after the delivery of feedback. Although a wealth of studies have tackled the relationship between FRN and feedback learning (for a review, see Walsh & Anderson, 2012), most studies relied on cross-trial averages to mirror learning processes on an electrophysiological level. As learning is a dynamic process, a trial-by-trial approach seems well suited to capture its dynamics. So far, only a few EEG studies have adopted such a parametric design by linking modeling-derived trial-by-trial learning signatures to single-trial amplitudes of ERPs (Hauser et al., 2014; Fischer & Ullsperger, 2013; Chase, Swainson, Durham, Benham, & Cools, 2011; Philiastides, Biele, Vavatzanidis, Kazzer, & Heekeren, 2010). Chase et al. (2011) reported a correlation of the FRN with model-free RPEs derived from a reinforcement learning model that only updates values of the chosen stimuli ("single-update model"). A recent study by Hauser et al. (2014) used a modified algorithm that simultaneously updates values of chosen and unchosen stimuli ("double-update model") and replicated the correlation between RPEs and the single-trial FRN. In addition, the authors studied the influence of unsigned PEs, a value-unspecific signal reflecting the unexpectedness of events, and found that unsigned PEs better accounted for their data. Hauser et al. (2014) concluded that the FRN codes surprise rather than a signed learning signal. Notably, both studies have not systematically separated uniquely explained variance by single-update versus double-update algorithms in the FRN.

Thus, the goal of this study was to identify correlates of abstract inference on the anticorrelated task structure in the FRN. To address this question, we recorded EEG data from 21 healthy participants while they performed counterfactual decision-making in a dynamically changing environment with anticorrelated reward probabilities. We evaluated the relation between single-trial FRN responses to feedback and RPEs, respectively. The following hypotheses were proposed: (1) The FRN signals a model-free RPE as suggested by prior studies, and (2) the FRN additionally signals values estimated through abstract inference on the anticorrelated task structure as derived from the difference between value estimates of double-update and single-update algorithms.

METHODS

Participants

Twenty-one healthy participants (age = 26.20 ± 3.49 years, range = 22–34 years, 11 women) were recruited from the Max Planck Institute's participant database. Participants received remuneration for participation in addition to the money won in the experimental task. The study was approved by the ethics committee of the University of Leipzig. All participants gave written informed consent to participate before beginning the study. One data set had to be excluded because of technical problems during EEG recording.

Task

During EEG acquisition, participants performed a probabilistic decision-making task in a dynamic environment with anticorrelated reward probabilities (Figure 1A) that requires flexible behavioral adaptation. In 160 trials, participants decided between two cards both showing an abstract geometric stimulus (maximum response time = 1500 msec). After the participant had chosen one stimulus using either the left or right button, the selected stimulus was highlighted and depicted for another 500 msec plus RT. After an ISI of 1500 msec (presentation of a fixation cross), one of two feedback stimuli was presented for 500 msec, indicating either a win (a 10 Euro-cent coin) or loss (a crossed 10 Euro-cent coin) of money. During the intertrial interval, a fixation cross was shown for a variable duration with a mean of ~2500 msec. Total trial duration was 6500 msec on average. The location (right vs. left side of the screen) where each of the stimuli was presented was randomized over trials. Crucially, reward contingencies were perfectly anticorrelated, that is, one of the stimuli was assigned a reward probability of 80% and a punishment probability of 20%, and vice versa for the other stimulus. This anticorrelation is essential because, if exploited by the agent, it enables the learner to infer the value of the unchosen stimulus, although no outcome is delivered for the unchosen option. Figure 2 provides a schematic of the idea of abstract inference on the task structure with respect to unchosen choice values ("double updating").

Reward contingencies remained stable for the first 55 trials and also for the last 35 trials. In between, reward contingencies changed four times: after 15 or 20 trials, each. This required participants to flexibly adapt their behavior and ensured constant learning (Figure 1B). Note that contingency reversals were predefined by the stimulation protocol and did not depend on participants' performance in the task. Before the experiment, participants were instructed that, depending on their choice, they could either win 10 Euro-cent or lose 10 Euro-cent per trial and that the total amount of money gained would be paid

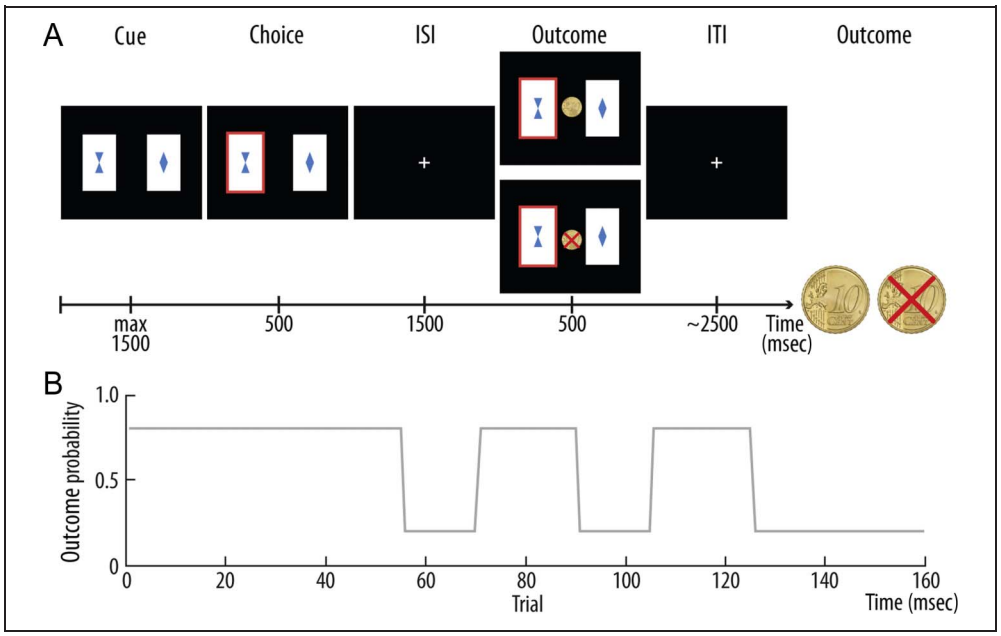

Figure 1. (A) Trial sequence of the serial reversal task. Participants were instructed to find the card with the superior chance of winning. One of the stimuli was assigned a reward probability of 80% and a punishment probability of 20% (vice versa for the other stimulus). Outcome stimuli were a 10-cent coin (in the case of a win) and a crossed 10-cent coin (in the case of a loss). (B) Reward contingencies over the course of the experiment. Reward contingencies remained stable for the first 55 trials and also for the last 35 trials. In between, reward contingencies changed four times. ITI = intertrial interval.

out at the end of the experiment. In addition, participants were informed that one of the two cards had a superior chance of winning money, but that this might change during the task. Participants became familiar with the task before the experiment by performing 20 training trials with different stimuli and without any reversal of reward contingencies. Distance between participant and computer screen was kept constant across all participants.

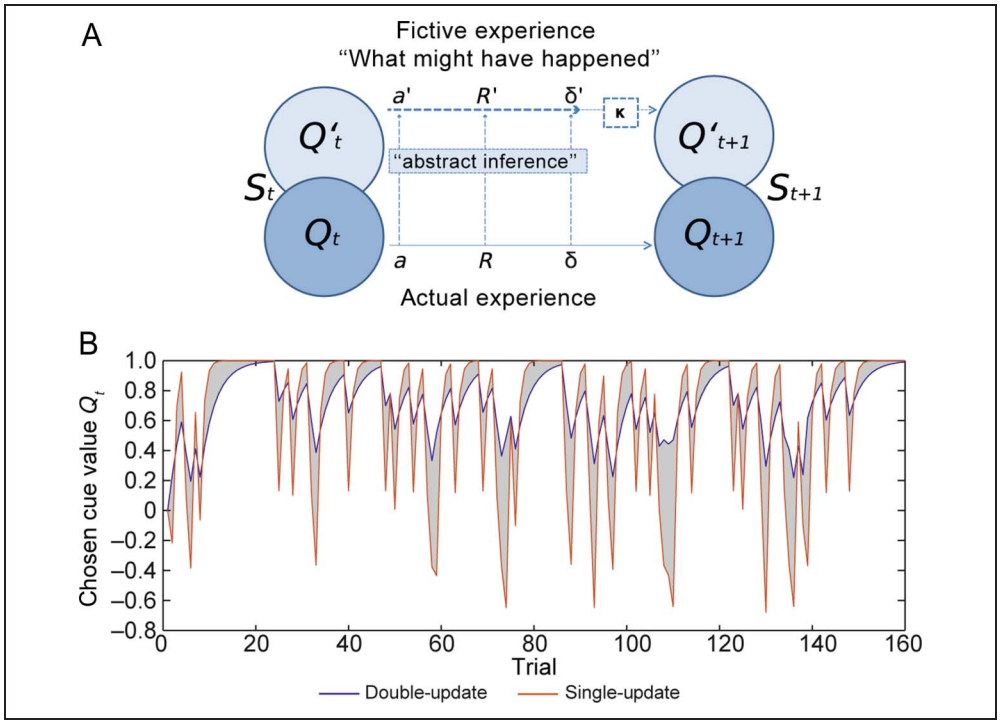

Figure 2. (A) At time t, an agent in state S_t passes to a new state S_{t+1} by the action a, observing the outcome R, which leads to the RPE δ, which is the difference between expected and actually gained outcomes. The agent updates the chosen value for the next trial, Q_{t+1}, accordingly. Although no outcome is delivered for the unchosen action, the agent can infer from the task structure, that is, the anticorrelation of the decision values, what might have happened (R') if he had chosen an alternative action a', resulting in a fictive PE δ′. Thus, by abstract inference about the counterfactual task structure and parallel to updating chosen values, the agent double-updates unchosen values Q'_{t+1}. Individuals might differ in their degree of abstract inference about the environmental structure. The individual degree of double-updating is therefore weighted by the parameter κ. (B) Effect of abstract inference, "double-updating," on chosen values. For one exemplary participant, values of the respective chosen value are plotted per trial, as a function of the two alternative control strategies: pure single-updating (κ = 0, neglecting "what might have happened," red) versus pure double-updating (κ = 1, full abstract inference on the task structure, blue). Hence, the difference of both (here, highlighted in gray) represents an estimate of the degree of abstract inference on the counterfactual task structure. We examine whether this difference in choice values with respect to abstract inference modifies the coding of the core teaching signal, the RPE δ for chosen values, and whether the FRN codes these components of abstract inference.

Computational Models of Learning

Implemented Models

The focus of this study was to examine trial-by-trial signatures of each individual's learning process in the electrophysiological recordings. Thus, the actually observed behavioral responses were analyzed in a computational modeling framework. Three different types of learning models were fitted to the data: (1) a single-update model that updates values only for the chosen stimulus, (2) a double-update model that updates values of chosen and unchosen stimuli with equal weight, and (3) a hybrid model that individually weights the degree of double-update learning.

Single-update Model

The single-update model updates a decision value $Q_{a,t}$ for the chosen stimulus via the RPE $\delta_{Q_{a,t}}$, which is defined as the difference between the received reward R_t and the expected reward for the chosen stimulus $Q_{a,t}$:

$$\delta_{Q_{a,t}} = R_t - Q_{a,t} \tag{1}$$

This teaching signal is then used to iteratively update chosen decision values trial-by-trial:

$$Q_{a,t+1} = Q_{a,t} + \alpha\delta_{Q_{a,t}} \tag{2}$$

Here, α represents a learning rate that weights the influence of reward RPEs on the updated values. This free parameter of the model has natural boundaries between 0 and 1. Note that this model neglects the anticorrelated structure of the task by simply updating decision values for chosen stimulus only, whereas the value of the unchosen stimulus $Q_{ua,t}$ remains unchanged:

$$Q_{ua,t+1} = Q_{ua,t} \tag{3}$$

Double-update Model

In a next step, we extended this model to more closely match the experimental environment by using abstract inference. In this task, an update of both decision values in each trial is advantageous because it takes into account the anticorrelated structure of the task. This double-update learning has been demonstrated to result in improved behavioral adaption to changing reward contingencies (Schlagenhauf et al., 2013; Li & Daw, 2011; Hampton et al., 2006). To mirror this strategy in our computational modeling approach, the unchosen decision values are updated using a different error signal. The prediction error (PE) for the double-update model is

$$\delta_{Q_{ua,t}} = -R_t - Q_{ua,t} \tag{4}$$

The same learning rate α is used to update the unchosen value:

$$Q_{ua,t+1} = Q_{ua,t} + \alpha\delta_{Q_{ua,t}} \tag{5}$$

We refer to this model as the double-update model, as it takes into account the values of the chosen and unchosen stimuli and thereby implements abstract inference on the anticorrelated task structure.

Hybrid Model

Note that Equation 5 gives the same weight to the update of unchosen decision values as to the chosen decision values. However, it is conceivable that the unchosen option is updated at a reduced rate of change when compared with the update of chosen values. Moreover, the degree of abstract inference may differ across individuals, as this is computationally more expensive and consequently may be limited depending on the specific situation that challenges the individual. This notion has been emphasized in healthy individuals (Radenbach et al., 2015; Eppinger, Walter, Heekeren, & Li, 2013; Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013; Wunderlich, Smittenaar, & Dolan, 2012; Daw, Gershman, Seymour, Dayan, & Dolan, 2011) and recently also in psychiatric disorders (Sebold et al., 2014; Voon et al., 2014). To account for this potential variability in employing abstract inference, we additionally tested a hybrid model including an individual degree of double-update learning. Therefore, the weighting parameter κ is introduced:

$$Q_{ua,t+1} = Q_{ua,t} + \kappa\alpha\delta_{Q_{ua,t}} \tag{6}$$

Note that Equations 1–5 refer to the special cases $\kappa = 0$ or $\kappa = 1$, respectively.

Observation Model

Finally, for all models, we transformed decisions into action probabilities by applying a softmax equation including the parameter β, which determines the stochasticity of the choices:

$$p(a,t) = \frac{\exp\left(\beta \times Q_{a,t}\right)}{\sum_{a'} \exp\left(\beta \times Q_{a',t}\right)} \tag{7}$$

where a' indicates all available choice options.

The sum of this probability $p_{(a,t)}$ over all trials (and participants, eventually) is the so-called negative log-likelihood, which is the probability of observing the data given the parameters of a model.

A priori Model Simulation

We simulated choice behavior 1000 times per model (single-update model: $\kappa = 0$, hybrid model: $\kappa = 0.5$, double-update model: $\kappa = 1$) by fixing all other free parameters to the mean estimate of an independent sample ($n = 35$)

performing the same task. The highest simulated mean of correct choices over the whole course of the experiment was observed for the hybrid model with $\kappa = 0.5$ (mean simulated correct choices [proportion]: single-update model = 0.81, hybrid model = 0.82, double-update model = 0.79). The advantage of the hybrid model, brought about by implementing abstract inference on the task structure ("double-updating"), becomes particularly prominent in trials with frequent reversals of reward contingencies (Trials 56–125; compare Figure 1B), which require flexible behavioral adaptation (mean simulated correct choices [proportion]: single-update model = 0.71, hybrid model = 0.77, double-update model = 0.77).

Model Fitting and Model Selection

Free parameters that have natural boundaries were fitted after transformation to a logistic (α) or exponential (β) distribution to render normally distributed parameter estimates. A maximum a posteriori estimate of each parameter for each participant was found by setting the prior distribution to the maximum likelihood given the data of all participants, and then expectation–maximization was used (for an in-depth description, compare Huys et al., 2011, with Huys et al., 2012). All modeling analyses were performed using MATLAB R2013a (The MathWorks, Inc., Natick, MA).

For all three models, we first report the negative log-likelihood and the Bayesian Information Criterion (BIC; Schwarz, 1978) based on the negative log-likelihood (Table 1). Note that the BIC includes a penalty term for the number of parameters in the model to account for the risk of overfitting. Second, the model evidence was approximated by integrating out the free parameters. The integral was approximated by sampling from the empirical prior distribution, and we therefore added the subscript "int" to the BIC (Table 1; see Huys et al., 2011, 2012). Third and reported in the text of the Results section, we subjected this integrated or marginalized likelihood to a random effects Bayesian model selection procedure (Stephan, Penny, Daunizeau, Moran, & Friston, 2009; spm_BMS contained in SPM8, www.fil.ion.ucl.ac.uk/spm/). After having identified the best-fitting model, we also verified that best-fitting parameters reproduce the observed behavior well by resimulating the task based on the inferred parameters.

Modeling-derived EEG Analysis

In line with previous reports, we regressed single-update RPEs against single-trial EEG data (Chase et al., 2011). To specifically address the question of whether or not the coding of RPEs in the FRN contains additional effects of abstract inference on the task structure as expressed in the double-update model, we computed a difference regressor (Daw et al., 2011), which was defined as the difference between the two error signals:

$$\delta_{a,t_{\text{difference}}} = \delta_{a,t_{\text{double-update}}} - \delta_{a,t_{\text{single-update}}} \quad (8)$$

Note that this regressor reflects differences of chosen decision values estimated by the single-update versus double-update algorithm (see Figure 2B for an illustration and Figure 4B for mean differences).

Electrophysiological Recording and ERP Extraction

Electrophysiological activity was recorded from 60 scalp positions according to the 10–20 EEG system (BrainAmp MR plus; Brain Products, Gilching, Germany).

Four additional ocular electrodes (vertical and horizontal EOG) were attached to monitor eye movements and blinks. EEG and EOG, referenced against the linked mastoids, were sampled at 2500 Hz (1-sec low cutoff, 250-Hz high cutoff, notch off). Electrode impedances were kept below 10 kΩ.

EEG data were preprocessed using EEGLAB 4.515 (Delorme & Makeig, 2004) and MATLAB R2013a. EEG data were down-sampled to 250 Hz and bandpass filtered between 0.5 and 46 Hz (butterworth filter, third order). Trials were segmented from −2 to 6 sec relative to the onset of the outcome stimulus. An independent component analysis (ICA; logistic Infomax ICA; Delorme & Makeig, 2004) was applied to decompose the multivariate EEG signal into statistically independent components. By two independent assessors, movement-related ICA sources and frontal sources with ocular artifacts, such as blinks and eye movements were visually identified by inspecting the scalp maps, time courses, and power spectra in all components, and were removed before back-projection of the remaining components onto the EEG channels. All EEG epochs were visually inspected before and after

Table 1. Model Comparison

	$-LL$	BIC	BIC_{int}	XP
Full hybrid model (κ as a free parameter)	849	751	808	0.9872
	Δ $-LL$ *Hybrid*	Δ *BIC Hybrid*	Δ BIC_{int} *Hybrid*	
Single-update model ($\kappa = 0$)	−34	26	12	0.0127
Double-update model ($\kappa = 1$)	−69	57	55	0.0001

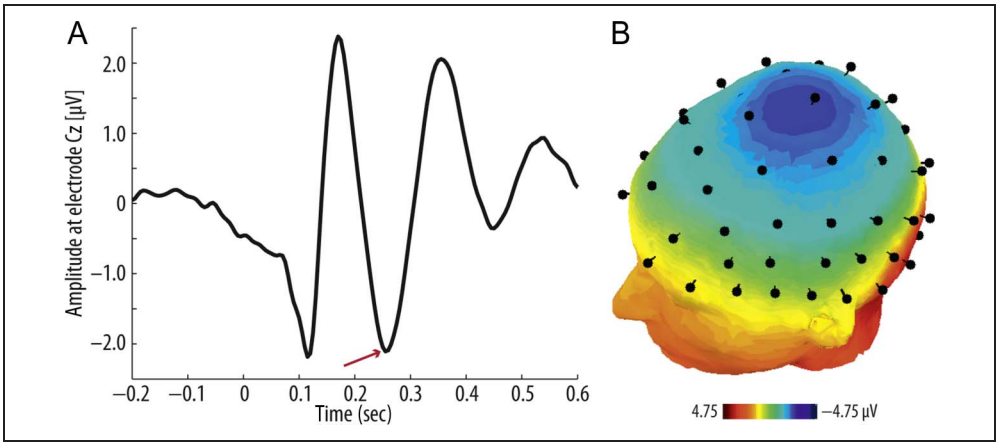

Figure 3. (A) Grand-averaged waveform of the FRN, revealing the FRN with its maximal amplitude 261 msec after feedback onset at the electrode Cz (here indicated by the arrow). (B) Topological distribution of the deflections at the time point of the average peak FRN.

ICA. Thereafter, data were re-referenced to the average (Lehmann & Skrandies, 1980).

For the analysis of the feedback-related electrophysiological responses, we identified the peak negativity at an a priori site of interest (Cz; according to Holroyd & Coles, 2002) and in a predefined time window of 200–300 msec after feedback onset (Hauser et al., 2014). We aimed to account for interindividual differences in ERP latencies by determining individual peak latencies. The individualized latency of the evoked components was derived from the individual participant's average FRN latency. This individualized time point was then used to extract single-trial amplitudes in all 160 trials. Figure 3B shows the averaged waveform at the electrode Cz and the topological distribution of the deflections at the time point of the average peak FRN across participants.

RESULTS

Behavioral Analyses

Task Performance

Participants chose the correct stimulus, that is, the stimulus with the higher reward probability, on average in 81% (SD = 6%) of all trials, indicating that participants understood and mastered the task appropriately. A mean of 0.25 (SD = 0.72) trials per participant had to be excluded for the computational modeling and single-trial EEG analysis because of missing responses.

Computational Modeling

Three different computational models were implemented to describe different ways of updating decision values during the learning process. The first, a single-update model, updates chosen values only. The second, a double-update model, additionally updates unchosen decision values. Compared with the second model, the third model quantified the degree of double-updating individually via the double-update weighting parameter κ. Bayesian model selection demonstrated that this hybrid model, a combination of both strategies quantified by the free parameter κ, explained the data best at the group level (Table 1). This suggests considerable interindividual variability in the extent to which individuals use double-updating regarding the unchosen option. Importantly, choice behavior of all participants was explained better than chance by the best-fitting model (mean explained choices = 78%, SD = 0.093%) considering the negative log-likelihood. Choices explained are in a similar range as in prior studies (Daw, 2011), suggesting that the winning model accounted well for the observed choice data. A simulation based on each individual's inferred parameters additionally showed that our model captured behavior well. The distribution of the best-fitting parameters and the negative log-likelihood is shown in Table 2.

The anticorrelated nature of the task leads to perfectly anticorrelated values of chosen and unchosen options. In line with the task structure, the correlation of chosen and unchosen cue values in the double-update model is approximately −1 (mean Pearson's r = −.989, SD = 0.010; note that it is slightly greater than −1 because of fitting initial cue values). If κ is meaningful in terms of explaining observed behavior, one would expect the correlation of chosen and unchosen cue values in the hybrid model to drop considerably and to be intermediate between single-update and double-update models. To test this, we also calculated mean correlation coefficients (Pearson's r) between chosen and unchosen values for the single-update model (which serves as a baseline for the correlation; as

Table 2. Distribution of Best-fitting Parameters (Hybrid Model with κ as a Free Parameter) and the Negative Log-likelihood

	β	*Initial Q*	α	κ	$-LL$
25th percentile	1.96	−.36	.53	.07	−54
Median	2.80	−.24	.57	.10	−33
75th percentile	3.80	−.07	.60	.17	−27

in this model, the unchosen value is not updated) and the hybrid model. Indeed, this revealed a mean $r = -.173$ ($SD = 0.173$) for the single-update model and a mean $r = -.416$ ($SD = 0.244$) for the hybrid model. An ANOVA model testing r derived from the three different models against each other confirmed that the correlation of cue values is significantly different in the described models ($F(2, 54) = 117$, $p < .001$). Post hoc t tests showed that all mean rs were significantly different from each other (all ts > 7.074, all ps $< .001$).

Correlation of FRN and Learning Signatures

Coding of Model-free RPEs in the FRN

In line with previous findings, we evaluated whether FRN amplitudes correlate with model-free RPEs. The relationship between the FRN amplitude and the magnitude of the model-free, single-update RPE was evaluated on a trial-by-trial basis using linear regression analysis. We found a significant positive correlation between this RPE and FRN amplitudes (mean slope of regression = 0.086, $SD = 0.116$, $t(19) = 3.31$, $p = .01$). In accordance with previous findings, this indicates that RPEs scale linearly with FRN amplitude, that is, that the FRN codes model-free RPEs

Coding of Abstract Double-update Inference in the FRN

In the next step, we directly aimed to determine the coding of single-update versus double-update components in the FRN amplitudes. Therefore, in addition to the model-free, single-update RPE regressor, we also entered a difference regressor between single-update and double-update RPEs into the multiple regression analysis. Note that the difference regressor as described in Equation 8 describes differences of decision values estimated by double-update minus the single-update algorithm, and thus represents the change in values uniquely associated with abstract double-update inference. We found a significant effect for both the single-update RPE regressor ($t(19) = 2.32, p = .032$) and this difference regressor (mean slope of regression = -0.085, $SD = 0.110$, $t(19) = -2.30$, $p = .033$). Note that the correlation with the model-free, single-update RPE is positive, indicating that RPEs scale with FRN amplitudes (see Figure 4A), whereas the correlation of FRN with the difference regressor is negative. The latter negative correlation reflects two key characteristics of abstract double-update inference:

1. By concurrently updating the unchosen choice option, double-updating maps the anticorrelated environment more precisely. This leads to differences in the size of RPEs from the double-update (RPE_{DU}) versus the single-update model (RPE_{SU}); as for the double-update learner, in certain cases, feedback is more predictable than for the single-update learner. This leads to a relatively attenuated RPE_{DU} in trials where RPE_{SU} has high absolute values. More detailed, in the case of unexpected punishments (defined by large negative RPEs), the RPE from the double-update model is less negative than the RPE from the single-update model; thus, the difference regressor is positive. On the contrary, in cases of relatively unexpected rewards (defined by large positive RPEs)—often after an agent has switched to the alternative option—the RPE from the double-update model is less positive than the RPE from the single-update model; consequently, the difference regressor becomes negative.
2. Double-update learning is smoother as recent events do not impact choice values as strongly as in single-update learning. Thus, after relatively expected punishments

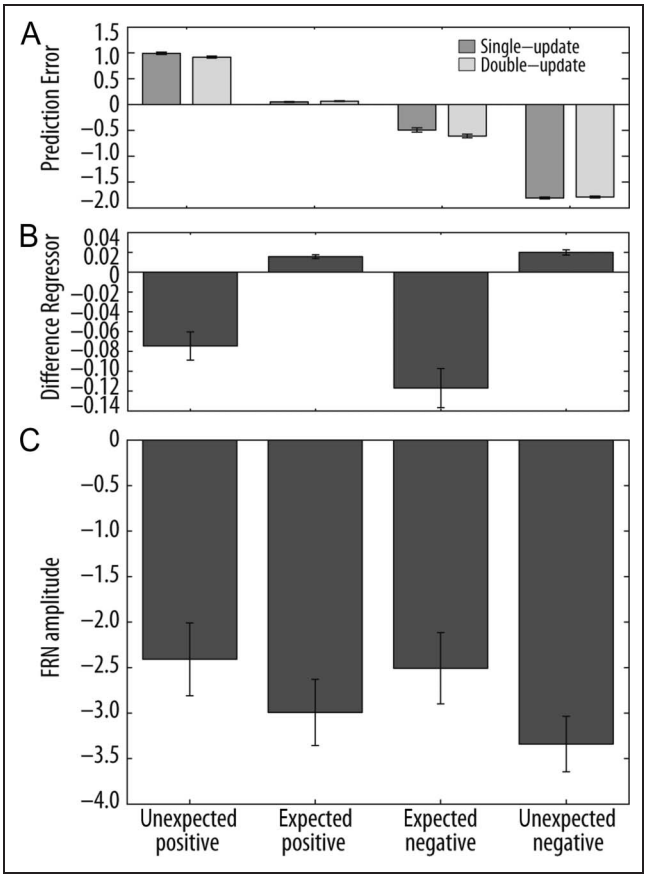

Figure 4. (A) Mean RPEs, derived from the single-update and double-update models, as a function of valence and expectedness. We plot mean RPEs of the 25th and 75th percentiles of each individual's range of RPEs. (B) Mean difference of chosen decision values estimated by the double-update and single-update models (Equation 8), that is, estimates of abstract "double-update" inference components in decision-making. Trial-by-trial differences were used as regressors to predict the electrophysiological signal. (C) Mean FRN amplitudes, plotted as a function of valence and expectedness of the feedback in the trial. FRN amplitudes were influenced by the interaction of valence and expectedness and showed a positive association with reward predicition errors. FRN amplitudes were furthermore negatively correlated with the difference regressor, indicating that the FRN codes abstract "double-update" inference components—the influence of "what might have happened"—in reward-guided decision-making.

(e.g., after a series of punishments, indicative of the necessity to switch to the alternative option), RPE_{SU} approaches zero faster, and RPE_{DU} is more negative than RPE_{SU}. This results in a negative difference regressor. Contrarily, in a rewarded trial, which has already been preceded by a series of rewards (e.g., after having learnt to stay with the better stimulus at one point in time), RPE_{DU} is numerically higher than RPE_{SU}. Thus, the difference regressor is positive in these cases. Figure 4 illustrates this description by plotting mean RPE_{SU} and RPE_{DU} as well as difference regressor and mean FRN amplitudes as a function of expectedness and valence of the feedback in the trial. Repeating the same analysis using the RPE derived from the double-update model instead of a difference regressor confirmed the results: Both predictors, the single-update ($t(19) = 3.77$, $p = .001$) as well as the double-update PE ($t(19) = -2.65$, $p = .016$), showed a significant effect.

Our observations suggest that, in addition to the model-free, single-update RPE, the difference regressor, which reflects the degree of abstract double-update inference, uniquely explains variance in the FRN amplitudes. These data indicate that, in addition to coding model-free, single-update RPEs, the FRN also signals values estimated through abstract inference on the anticorrelated task structure.

Influence of Signed versus Unsigned RPEs

It had been suggested that, rather than a reinforcement learning RPE signal, the FRN might reflect a surprise signal. Unsigned model-derived PEs ($|PEs|$) represent such a surprise signal. For the sake of replication, we repeated the analyses reported by Hauser et al. (2014) and entered signed PEs derived from the double-update model as well as their unsigned values $|PEs|$ in a multiple regression analysis. The resulting two beta weights were then analyzed with a t test (Holmes & Friston, 1998). Note that, although signed and unsigned PE values are correlated, this analysis only accounts for uniquely explained variance, and betas derived from this analysis are not pseudoeffects of the correlated measure. We found a significant effect of the signed RPEs on the single-trial amplitude (mean slope of regression = 0.082, $SD = 0.114$, $t(19) = 3.418$, $p = .023$), whereas there was a trend for the unsigned RPEs (mean slope of regression = 0.019, $SD = 0.090$, $t(19) = 1.877$, $p = .076$). Similar findings were obtained when binning small RPEs and large RPEs for wins and losses separately by identifying the 25th and 75th percentiles of each individual's range of RPEs and testing the effect of RPE size and feedback valence on FRN amplitudes using a repeated-measures ANOVA. We found a significant interaction effect of valence and RPE size ($F = 10.68$, $p = .004$; see Figure 4C), whereas no significant main effect of size and valence was observed (all ps $> .10$, all Fs ≤ 2.81). We conclude that signed double-update RPEs uniquely explain variance in the FRN trial-by-trial amplitudes. Thus, we argue that the FRN codes learning signals rather than mere surprise (Hauser et al., 2014).

DISCUSSION

Although it is essential for an agent to learn from observed outcomes emerging as a consequence of actual choice, hypothetical inference on "what might have happened" is thought to additionally guide decision-making and improve behavioral adaptation. In this study, we could identify separate contributions of single-update versus double-update learning, the latter reflecting abstract inference on the task structure, by focusing on unique variances explained by the difference between their value estimates in the FRN. Thereby, we demonstrate that the FRN codes the influence of both a model-free, single-update RPE and additional components related to abstract inference about the anticorrelated task structure.

Revisiting the Role of the FRN in Reinforcement Learning

For the investigation of electrophysiological signatures of reinforcement learning, a candidate deflection is the FRN. An influential theory suggests that the FRN is a neural signature of model-free, single-update RPE processing (Holroyd & Coles, 2002). Studies using cross-trial averages and recently also parametric analyses based on computational modeling have partially confirmed these theoretical claims (for a review, see Walsh & Anderson, 2012). Our findings are in line with and extend Holroyd and Coles' (2002) seminal theory: We argue that the FRN in fact mirrors model-free learning signals but additionally codes influences of inferred stimulus values deduced via abstract inference about the task structure.

Inference about Alternative Outcomes as a Feature of Flexible Behavioral Adaptation

Learning from experiential, observed outcomes versus inferred outcomes based on abstract inference has been related to the distinction between model-free and model-based control of behavior (Lucantonio et al., 2012; Bromberg-Martin et al., 2011). Model-free control is driven by rewards achieved in the past and is therefore retrospective and reflexive. By contrast, model-based behavior as the deliberative, prospective mode of control relies on an internal representation of the environment and allows forward planning of future actions based on their potential outcomes. Consequently, model-based control is computationally more expensive but enables individuals to rapidly adapt their behavior in a dynamically changing environment (Dolan & Dayan, 2013; Daw, Niv, & Dayan, 2005). In this study, the double-update model modifies model-free learning signals by incorporating the anticorrelated task structure, which leads to more successful

behavioral adaptation in a dynamic environment. It is therefore conceivable that abstract double-update inference is associated with the model-based system, as the model-free system is by definition blind toward the environmental structure. Our result that the FRN codes both model-free single-update RPEs and additional components reflecting abstract inference on the anticorrelated task structure fits neatly to the notion of a common architecture for the human control systems over decision-making with ubiquitous higher-order model-based influences in neural reward processors (Doll, Simon, & Daw, 2012). In line with previous studies, this goes against dual system accounts of isolated model-free versus model-based control. However, an alternative explanation might include that the abstract double-update inference as observed in this study does not arise from a full model-based system but rather temporal difference learning about the relationship of the choice options (Doll et al., 2012; Wimmer, Daw, & Shohamy, 2012; Shohamy & Wagner, 2008). Although there is no unique formulation of model-free and model-based control (Dolan & Dayan, 2013) and tasks may differ in which aspects of the system they capture, another approach of dissociating model-based versus model-free decision-making is to use sequential decision tasks (Daw et al., 2011; Glascher, Daw, Dayan, & O'Doherty, 2010). In sequential decision-making, the model-based learner acquires knowledge about the task-immanent transition structure and uses this to evaluate decision options. This task differs from the experiment used here by capturing a more complex learning environment and thereby offers the possibility of investigating one important feature of model-based control, namely, inferring action values by a learnt cognitive sequential model of the consequences of one's actions (Doll et al., 2012). Our findings encourage further electrophysiological investigations of different aspects of behavioral control, for example, in more complex environments via the application of sequential learning tasks.

A point worth considering is that, although the task structure uses perfectly anticorrelated reward probabilities, the degree of double-updating (as given by the parameter κ) is relatively low in the sample at hand and thus learners' chosen and unchosen values are only moderately correlated. This, together with the superiority of the hybrid model revealed by model selection, has two interesting implications: (1) a learner indeed updates the unchosen option; however, (2) updating the unchosen option happens only to a moderate degree in this task and less than updating the actually chosen option. This is the case despite the perfect anticorrelation of the true reward probabilities. As updating of the unchosen option depends on updating the chosen option, it is likely to be a more implicit process. Note that the anticorrelation of reward probabilities was not instructed before the experiment. It is thus conceivable that concurrent updating of both choice options might be computationally more demanding than limiting the update process only to one of the two stimuli. It might thus be plausible that, in a task such as the one used here, with phases of high stability in which double-updating is not necessary, the learner engages in the additionally demanding process to a reduced degree. A similar mechanism has been proposed for the model-free/model-based dichotomy, where the model-based system is believed to come into play only in environmental conditions that require flexibility (Keramati, Dezfouli, & Piray, 2011; Daw et al., 2005). This has important implications for future studies, which might investigate the following research questions: (1) Are there interindividual differences in how much individuals engage in double-updating of alternative options? Might reduced updating of unchosen options be associated with certain psychiatric states, characterized by maladaptive decisions? (2) Is double-updating a dynamic process, which covaries with the uncertainty of the environment?

Potential Generators of the FRN and Neural Correlates of Behavioral Control

Compared with methods such as fMRI, EEG is limited in tracking signals from deep subcortical structures that have been postulated to play key roles in reinforcement learning, such as the striatum. However, the plausible claim has been made that local field potentials are modulated by afferent midbrain PE signals (Talmi, Fuentemilla, Litvak, Duzel, & Dolan, 2012). Possible generator regions of the FRN are a matter of debate. Notably, the FRN is measured over the medial frontal cortex, a region that has been implicated in the coding of model-free as well as model-based RPE signals (Daw et al., 2011). Interestingly, studies using EEG source localization discuss the origin of the FRN in the striatum (Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011; Foti, Weinberg, Dien, & Hajcak, 2011), a brain structure that has also been shown to be involved in the processing of both model-free and model-based learning signals (Daw et al., 2011). Recently, a combined EEG–fMRI study likewise adopted a single-trial approach to track coupling of feedback signals in hemodynamic and electrophysiological responses (Becker, Nitsch, Miltner, & Straube, 2014). Their data imply contributions of multiple frontal midline generators to the FRN signal. Notably, hemodynamic activity in the medial pFC and the striatum was also coupled to the magnitude of electrophysiological FRN responses.

FRN and Signed versus Unsigned RPEs

Whether the FRN codes signed or unsigned RPEs is a matter of ongoing debate (Cavanagh & Frank, 2014; Hauser et al., 2014; Ullsperger, Fischer, Nigbur, & Endrass, 2014; Talmi, Atkinson, & El-Deredy, 2013; Alexander & Brown, 2011). A previous modeling-based study found a correlation of FRN with unsigned RPEs. The authors conclude that the FRN rather reflects salience coding than

learning (Hauser et al., 2014). Our findings contribute to this discussion by showing that the FRN is explained by signed RPEs and also reflects abstract higher-order components. This points toward a role for the FRN in learning beyond coding of expectedness or salience only. However, albeit on a trend level only, we also found a correlation between the FRN and unsigned PEs. Our findings are in line with a recent interpretation, which discusses contributions of signed versus unsigned RPEs to the FRN (Ullsperger et al., 2014): The authors argue that the presence of both signed and unsigned RPEs in the FRN is plausible because unsigned RPEs may, beyond signed RPEs, play a particular role in learning and behavioral adaptation. The authors suggest that surprise signals can be used to modify a weighting factor (such as learning rate or volatility of the environment) of signed RPEs. Although our findings corroborate this unifying notion, we moreover suggest a closely related interpretation of the association of unsigned PEs with the FRN. The absolute value of the model-free RPE signal has been claimed to function as information on the reliability of the model-free system (Lee, Shimojo, & O'Doherty, 2014; Roesch, Esber, Li, Daw, & Schoenbaum, 2012). Such reliability signals are thought to be used by an arbitration mechanism, which allocates the degree of control exerted by one of the systems at a given point in time. On the basis of the findings by Hauser et al. (2014), and also the trendwise correlation observed in our data, we suggest that components of the FRN additionally code the reliability of the model-free system and may thereby also reflect an electrophysiological signature of this arbitration process. Specific modeling strategies are warranted to address this question (e.g., Lee et al., 2014; Roesch et al., 2012; Li, Schiller, Schoenbaum, Phelps, & Daw, 2011).

Limitations

It has been argued that limitations inherent to the ERP methodology render conclusions about the role of the FRN in reinforcement learning opaque (Cohen, Mimes, & van de Vijver, 2011; Cavanagh, Frank, Klein, & Allen, 2010). Our findings suggest that, by adopting a modeling-derived parametric approach, FRN accounts can contribute to a more profound understanding of electrophysiological correlates of human decision-making processes. Building on that, we believe that, for future electrophysiological studies in the framework of behavioral control, it seems promising to additionally take dynamic changes in systems level oscillatory synchronization into account (Cavanagh & Frank, 2014; Cohen et al., 2011; Cavanagh et al., 2010).

In conclusion, our findings provide an electrophysiological correlate of incorporating abstract inference into the decision-making process. Reduced neural tracking of PEs (Parvaz et al., 2015; Tanabe et al., 2014), disturbed mechanisms of inference (Huys, Guitart-Masip, Dolan, & Dayan, 2015; Lucantonio, Takahashi, et al., 2014; Lucantonio et al., 2012), and altered behavioral control—for example, an imbalance between model-based and model-free control—are suggested to have psychopathological implications (Reiter, Deserno, Wilbertz, Heinze, & Schlagenhauf, 2016; Huys & Petzschner, 2015; Gillan & Robbins, 2014; Lucantonio, Caprioli, & Schoenbaum, 2014; Deserno, Boehme, Heinz, & Schlagenhauf, 2013; Dolan & Dayan, 2013). For instance, patients experiencing disorders characterized by failure in behavioral adaptation, such as, addiction or obsessive compulsive disorder, have been reported to show reduced model-based learning (Sebold et al., 2014; Voon et al., 2014). As EEG, in comparison with fMRI, is advantageous with regard to feasibility, we offer new means of studying these processes in patient populations characterized by aberrant reinforcement learning mechanisms.

Acknowledgments

The authors thank T. A. Klein for helpful comments on the experimental design and on an earlier version of the manuscript, S. Stasch for assistance in EEG data acquisition, H. Schmidt-Dudersted for her help in designing the figures, and A. Calder and E. Kelly for proofreading.

Reprint requests should be sent to Andrea M. F. Reiter, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany, or via e-mail: reiter@cbs.mpg.de.

REFERENCES

Abe, H., & Lee, D. (2011). Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron, 70,* 731–741.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14,* 1338–1344.

Becker, M. P. I., Nitsch, A. M., Miltner, W. H. R., & Straube, T. (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *Journal of Neuroscience, 34,* 3005–3012.

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biology, 9,* e1001093.

Boorman, E. D., Behrens, T. E., Woolrich, M. W., & Rushworth, M. F. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron, 62,* 733–743.

Bromberg-Martin, E. S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. *Journal of Neurophysiology, 104,* 1068–1076.

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *Neuroimage, 57,* 1608–1616.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18,* 414–421.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage, 49,* 3198–3209.

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience, 23,* 936–946.

Cohen, M. X., Mimes, K. A., & van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends in Cognitive Sciences, 15,* 558–566.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In *Decision making, affect, and learning: Attention and performance XXIII* (Vol. 23, pp. 3–38). Oxford, UK: Oxford University Press.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204–1215.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Deserno, L., Boehme, R., Heinz, A., & Schlagenhauf, F. (2013). Reinforcement learning and dopamine in schizophrenia: Dimensions of symptoms or specific features of a disease group? *Frontiers in Psychiatry, 4,* 172.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80,* 312–325.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology, 22,* 1075–1081.

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience, 7,* 253.

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron, 79,* 1243–1255.

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping, 32,* 2207–2216.

Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive compulsive disorder. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 369,* 20130475.

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

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex, 19,* 483–495.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26,* 8360–8367.

Hauser, T. U., Iannaccone, R., Stampfli, P., Drechsler, R., Brandeis, D., Walitza, S., et al. (2014). The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. *Neuroimage, 84,* 159–168.

Holmes, A. P., & Friston, K. J. (1998). Generalisability, random effects & population inference. *Neuroimage, 7,* S754.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Huys, Q. J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Computational Biology, 7,* e1002028.

Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology, 8,* e1002410.

Huys, Q. J., Guitart-Masip, M., Dolan, R. J., & Dayan, P. (2015). Decision-theoretic psychiatry. *Annual Review of Neuroscience, 38,* 1–23.

Huys, Q. J., & Petzschner, F. H. (2015). Failure modes of the will: From goals to habits to compulsions?. *American Journal of Psychiatry, 172,* 216–218.

Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology, 7,* e1002055.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron, 81,* 687–699.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48,* 609–621.

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience, 31,* 5504–5511.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience, 14,* 1250–1252.

Lucantonio, F., Caprioli, D., & Schoenbaum, G. (2014). Transition from "model-based" to "model-free" behavioral control in addiction: Involvement of the orbitofrontal cortex and dorsolateral striatum. *Neuropharmacology, 76,* 407–415.

Lucantonio, F., Stalnaker, T. A., Shaham, Y., Niv, Y., & Schoenbaum, G. (2012). The impact of orbitofrontal dysfunction on cocaine addiction. *Nature Neuroscience, 15,* 358–366.

Lucantonio, F., Takahashi, Y. K., Hoffman, A. F., Chang, C. Y., Bali-Chaudhary, S., Shaham, Y., et al. (2014). Orbitofrontal activation restores insight lost after cocaine use. *Nature Neuroscience, 17,* 1092–1099.

Parvaz, M. A., Konova, A. B., Proudfit, G. H., Dunning, J. P., Malaker, P., Moeller, S. J., et al. (2015). Impaired neural response to negative prediction errors in cocaine addiction. *Journal of Neuroscience, 35,* 1872–1879.

Philiastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. *Neuroimage, 53,* 221–232.

Radenbach, C., Reiter, A. M., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H. J., et al. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology, 53,* 268–280.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology, 20,* 262–270.

Reiter, A. M. F., Deserno, L., Wilbertz, T., Heinze, H. J., & Schlagenhauf, F. (2016). Risk factors for addiction and their association with model-based behavioral control. *Frontiers in Behavioral Neuroscience, 10*.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *European Journal of Neuroscience, 35,* 1190–1200.

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron, 70,* 1054–1069.
Schlagenhauf, F., Rapp, M. A., Huys, Q. J., Beck, A., Wustenberg, T., Deserno, L., et al. (2013). Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Human Brain Mapping, 34,* 1490–1499.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.
Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., et al. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology, 70,* 122–131.
Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron, 60,* 378–389.
Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron, 80,* 914–919.
Sokol-Hessner, P., Hutcherson, C., Hare, T., & Rangel, A. (2012). Decision value computation in DLPFC and VMPFC adjusts to the available decision time. *European Journal of Neuroscience, 35,* 1065–1074.
Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage, 46,* 1004–1017.
Takahashi, Y. K., Chang, C. Y., Lucantonio, F., Haney, R. Z., Berg, B. A., Yau, H. J., et al. (2013). Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron, 80,* 507–518.
Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience, 33,* 8264–8269.
Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., & Dolan, R. J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. *Neuroimage, 59,* 635–645.
Tanabe, J., Reynolds, J., Krmpotich, T., Claus, E., Thompson, L. L., Du, Y. P., et al. (2014). Reduced neural tracking of prediction error in substance-dependent individuals. *American Journal of Psychiatry, 170,* 1356–1363.
Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences, 18,* 259–267.
Voon, V., Derbyshire, K., Ruck, C., Irvine, M. A., Worbe, Y., Enander, J., et al. (2014). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry, 20,* 345–352.
Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews, 36,* 1870–1884.
Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience, 35,* 1092–1104.
Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron, 75,* 418–424.