Neuron
Article

Real and Fictive Outcomes Are Processed Differently but Converge on a Common Adaptive Mechanism

Adrian G. Fischer[1,3,*] and Markus Ullsperger[1,2,3,4]
[1]Otto-von-Guericke University, Institute for Neuropsychology, 39106 Magdeburg, Germany
[2]Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6525 HR Nijmegen, the Netherlands
[3]Max Planck Institute for Neurological Research, 50931 Cologne, Germany
[4]Center for Behavioral Brain Sciences, 39106 Magdeburg, Germany
*Correspondence: adrian.fischer@ovgu.de
http://dx.doi.org/10.1016/j.neuron.2013.07.006

SUMMARY

The ability to learn not only from experienced but also from merely fictive outcomes without direct rewarding or punishing consequences should improve learning and resulting value-guided choice. Using an instrumental learning task in combination with multiple single-trial regression of predictions derived from a computational reinforcement-learning model on human EEG, we found an early temporospatial double dissociation in the processing of fictive and real feedback. Thereafter, real and fictive feedback processing converged at a common final path, reflected in parietal EEG activity that was predictive of future choices. In the choice phase, similar parietal EEG activity related to certainty of the impending response was predictive for the decision on the next trial as well. These parietal EEG effects may reflect a common adaptive cortical mechanism of updating or strengthening of stimulus values by integrating outcomes, learning rate, and certainty, which is active during both decision making and evaluation. Neuronal processing of real (rewarding, punishing) and fictive action outcomes (which would have happened had one acted differently) differs for 400 ms and then converges on a common adaptive mechanism driving future decision making and learning.

INTRODUCTION

Wouldn't it be nice to know what would have happened if you had chosen differently? Imagine driving on a highway toward a traffic jam faced with two choices: bypass the highway or wait in the hold up. Neither of the cases provides information about which decision really yields the better result. On the other hand, when choosing between two lanes in a traffic jam, you will always notice the progress you are making in your lane and the progress you could have been making in the other lane. Both humans (Burke et al., 2010) and monkeys (Subiaul et al., 2004) share the ability to learn complex rules and values from watching the actions of other conspecifics—termed vicarious or observational learning. This capability provides evolutionary benefits by reducing trial and error learning costs and can be speculated to be the progenitor of more abstract, counterfactual reasoning in humans. In reinforcement-learning models, it has been theorized that learning can be based on results from unchosen options as well (Sutton and Barto, 1998). Although the neural implementation of counterfactual learning recently sparked considerable interest (Boorman et al., 2011), little is known about its exact timing—particularly with regard to the processing of fictive prediction errors (PEs) (Chiu et al., 2008; Lohrenz et al., 2007) and their neural realization in the absence of other actors (de Bruijn et al., 2009).

To study the temporospatial evolution of cortical brain activity during learning from real and fictive outcomes and behavioral choice based on the learned stimulus values, we used a probabilistic reinforcement-learning task while recording electroencephalogram (EEG). Subjects decided to either choose or avoid gambling following one centrally presented stimulus in every trial (Figure 1A). A chosen gamble resulted in a monetary gain or loss, depending on the reward contingency associated with that stimulus. In choosing not to gamble, subjects avoided financial consequences, yet still observed what would have happened if they had chosen to gamble. Although neither directly rewarding or punishing, fictive outcomes can be used in the same way as real outcomes to update learned estimated values of given stimuli and determine whether behavioral adjustments are needed. Notably, the subjective valence of the feedback reverses after avoiding a gamble: a fictive and thus foregone reward (reflected in a positive PE in our computational reinforcement learning model, see Experimental Procedures and further below) is unfavorable, and a fictive and thus avoided loss (reflected in a negative PE) is favorable (Figure 1B). *Good*, *bad*, and *neutral* stimuli were presented; their valence was reflected in reward probabilities above, below, or at chance level, respectively. By learning which symbols to choose and which to avoid, subjects could maximize their earnings.

RESULTS AND DISCUSSION

Behavior and Computational Model

Subjects learned avoiding bad and choosing good stimuli comparably well: we observed no difference in the absolute

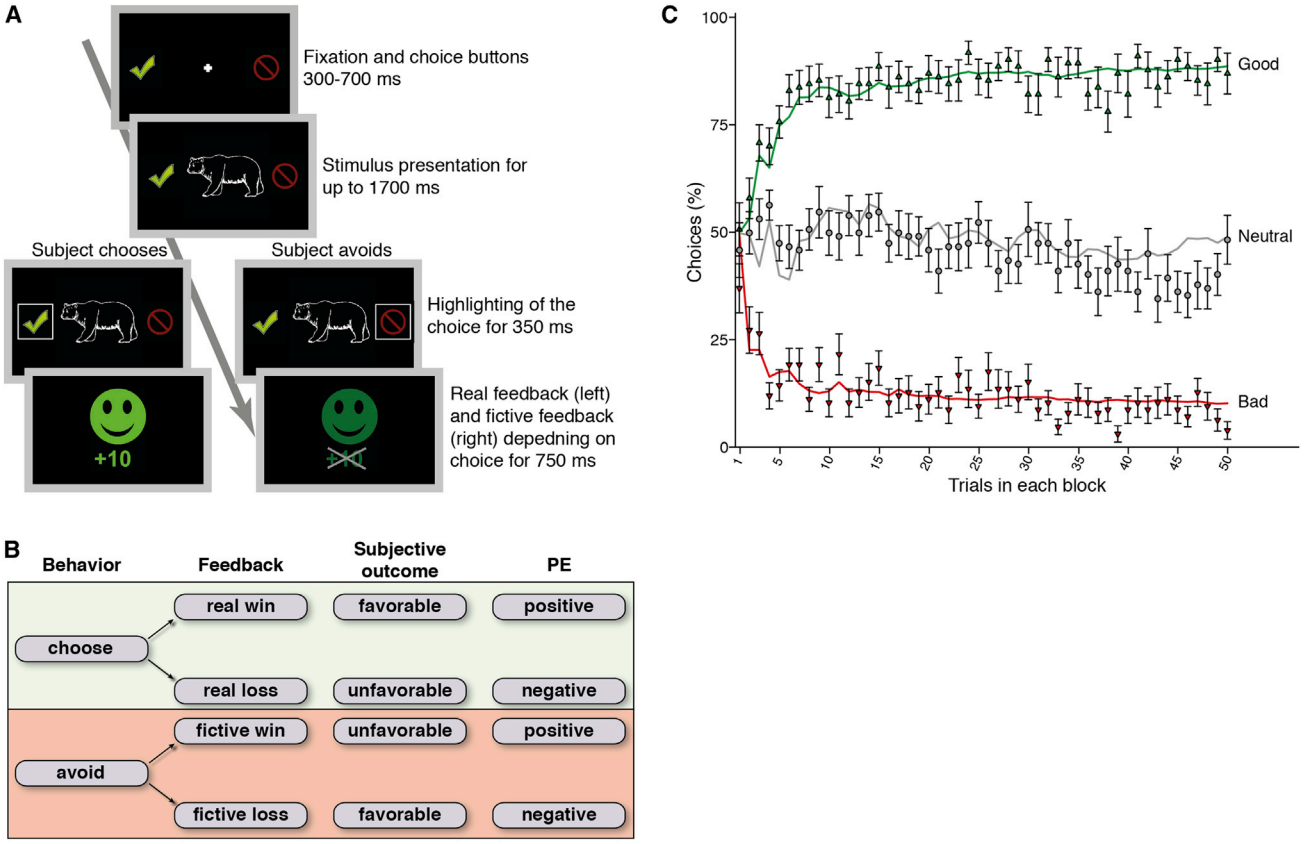

Figure 1. Experimental Design, Modeled and Observed Behavior

(A) Time course of the learning task. Choosing to gamble following stimulus presentation leads to real feedback consisting of a win or loss of 0.10€. Avoiding the gamble leads to fictive feedback without financial consequences but still provides information about the outcome of the trial.

(B) Task structure separated by subjects' choices. Note that the sign of the PE reverses in relation to the subjective outcome depending on the choice made.

(C) Modeled and observed behavior. Learning curves for empiric behavior of all subjects (symbols, ±SE) and predictions of the computational model (solid lines in the same color) for *good*, *bad*, and *neutral* symbols. Learning curves were comparable in both conditions and approached asymptotically toward their final levels. See Figure S1.

number of correct decisions following good compared to bad stimuli (t_{30} = 1.31, p = 0.20). Additionally, median reaction times did not differ between conditions (t30 = 0.43, p = 0.67). Learning of choice behavior for good and bad stimuli followed a logarithmic curve approaching an asymptote reflecting the probabilistic outcome of the respective stimuli (Figure 1C). This supports the notion that the weight of reward PEs in value updating decreases in an exponential fashion.

To derive single-trial estimates of individual PEs and subjective stimulus values, we fit a Q-learning model (see Experimental Procedures) (Jocham et al., 2009; Sutton and Barto, 1998; Watkins and Dayan, 1992) to each subject's sequence of choices. To account for the observed decrease in learning, we implemented an exponentially decreasing half-life time as a free model parameter that reduces the learning rate in later trials providing single-trial estimates of the learning rate (α_t). Maximum likelihood estimated (MLE) learning parameters of the model did not differ for learning from real and fictive outcomes (Table 1), indicating that subjects could utilize both sources of information with similar efficiency. This is also supported by the fact that sensitivity to misleading probabilistic feedback did not differ significantly between real and fictive conditions (Supplemental Information available online). MLEs of the half-life time indicated an average decrease of α_t of more than 90% in both conditions per block. Additionally, negative log-likelihood ($-LL$) did not differ when compared between good and bad stimuli.

Early Dissociation of Feedback Processing

Submitting feedback-locked EEG epochs to multiple robust regression analysis (Cohen and Cavanagh, 2011; O'Leary, 1990; Rousselet et al., 2008) revealed a double dissociation of cortical PE correlates between real and fictive outcomes in the first 400 ms following feedback. Intriguingly, the first significant covariation of feedback-locked EEG activity with PEs was found exclusively for fictive outcomes: a negative early occipital effect occurred 192–238 ms after feedback (Figure 2A and Movie S1) and was localized to extrastriate visual and posteromedial cortex (PMC; Figure S2A). In contrast, only real outcomes were associated with a somewhat later positive early PE effect spanning from 236–294 ms and a subsequent negative midlatency frontal

Table 1. MLE Parameter and Model Fit

Parameter	Real (SE)	Fictive (SE)	p for Difference
α_1	0.484 (0.068)	0.421 (0.065)	0.497
Hl	9.781 (2.606)	13.467 (3.309)	0.403
β	10.356	10.356	–

Neither the initial learning rate α_1 nor the half-life time *Hl* differed significantly when estimated for real and fictive outcomes separately. The sensitivity parameter β was kept the same for both conditions to ensure that models were comparable. $-LL$ showed significantly lower values for both *good* (69.363) and *bad* (63.284) compared to *neutral* stimuli (116.641; difference: $t_{30} > 8.7$, $p < 10^{-7}$ for both), while no difference between *good* and *bad* was observed ($t_{30} = 1.23$, $p = 0.229$). Therefore, our model was equally effective in describing subjects' behavior for *good* and *bad* symbols, yet less effective for *neutral* ones.

PE covariation at 336–430 ms, which in the averaged event-related potentials (ERPs) give rise to the feedback-related negativity (FRN) and P3a components, respectively (Figure 3). Direct contrasts between both conditions showed significant differences at electrode Oz during the time window of the occipital PE effect (peak $t_{30} = -4.18$, 204 ms, $p < 0.0005$) and at electrode FCz during FRN (peak $t_{30} = 4.95$, 284 ms, $p < 10^{-4}$), as well as P3a time windows (peak $t_{30} = -7.95$, 394 ms, $p < 10^{-8}$) (Figure 4B). The temporospatial double dissociation in early processing of real and fictive feedback was statistically confirmed by a triple interaction of the factors electrode, time window, and condition in an ANOVA on the average regression weights of the early PE effects in significant time windows (190–240 ms and 250–300 ms, for fictive and real feedback, respectively) at the most significant electrodes (Oz and FCz, for fictive and real feedback, respectively).

The FRN is usually found on unfavorable outcomes that violate expectancies (Gehring and Willoughby, 2002; Miltner et al., 1997). Our findings are consistent with an influential theory proposing that the FRN reflects PE signals (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004). The negative polarity of the FRN is in accordance with a positive covariation, as unfavorable real outcomes cause negative PE values. It has been consistently localized to the posterior medial frontal cortex (pMFC) (Gehring and Willoughby, 2002; Gruendler et al., 2011; Miltner et al., 1997), which has been supported by fMRI findings on feedback processing (Ridderinkhof et al., 2004; Ullsperger and von Cramon, 2003). The subsequent pronounced negative midlatency frontal PE effect fits well with theories relating the P3a to the recruitment of attention (Polich, 2007), which is here caused by negative PEs leading to a negative covariation by instigating increased P3a amplitudes. Exploratory localization analysis suggests a source network in cingulate gyrus and orbitofrontal cortices (Figure S2B).

In stark contrast to the real feedback condition associated with the well-known pattern reflecting FRN and P3a, following fictive feedback, these early and midlatency frontal PE effects were conspicuously absent; the average ERP waveforms showed merely a small negative deflection in the FRN time window that was unmodulated by learning parameters (Figures 3 and 4A). Feedback-related pMFC activity has been proposed to reflect action value updating (Amiez et al., 2006; Jocham et al., 2009; Kennerley et al., 2006; Walton et al., 2004). This suggests that a previous action is required in order to involve pMFC in the rapid processing of expectancy violations. The absence of an FRN-like PE effect on fictive outcomes could be explained in two ways: avoiding a stimulus is interpreted as abstaining from an action, or the neutral monetary outcome does not yield the necessary PE signal required for credit assignment to avoiding. The latter explanation seems very unlikely as other cortical PE correlates were found for fictive outcomes and MLE learning parameters in our task do not differ between conditions. It is also unlikely that the missing FRN results from reduced expectancy of and attention to fictive outcomes, because behavioral and modeling data as well as later EEG effects (see below) suggest similar utilization of fictive and real feedback. The absence of the FRN on fictive outcomes seems at odds with studies reporting FRN-like EEG deflections and pMFC activity on observed errors and feedback to others' actions (de Bruijn et al., 2009; van Schie et al., 2004; Yu and Zhou, 2006). Yet, in contrast to abstaining from choosing a stimulus in our experiment, observing actions could also lead to action simulation effects in motor-related areas via mirror systems (Rizzolatti et al., 2001)—permitting an update of action values. Taken together, it appears most likely that for motor-related areas, such as the pMFC, avoiding a stimulus in our learning task is equivalent to not performing any motor action.

However, the absence of the FRN and P3a modulation by fictive PEs does by no means indicate that outcomes are not processed in the fictive task condition. In sharp contrast to real feedback, we observed an early occipital PE-related EEG modulation following fictive feedbacks that even precedes the FRN time window, which has previously been interpreted as the fastest cortical correlate of feedback processing (Gehring and Willoughby, 2002; Philiastides et al., 2010). Its very short latency and localization to extrastriate visual areas and PMC (Figure S2A) seem to suggest that fictive outcomes engage a specific mechanism that might ease counterfactual learning. Although EEG does not allow precise localization, the found source fits well with findings from fMRI studies in which PMC has been associated with tracking values and PE signals of alternative unchosen options coding a counterfactual PE (Boorman et al., 2011). In monkeys (Leichnetz, 2001) and humans (Mars et al., 2011), the PMC is intensely interconnected with the more lateral part of the parietal cortex that has been shown to code fictive PE signals defined as the value difference between outcomes that could have been attained by optimal investments and actually attained outcomes (Chiu et al., 2008; Lohrenz et al., 2007). Furthermore, afferent projections from the basal forebrain as well as reciprocal projections with the anterior cingulate cortex shown in macaques (Parvizi et al., 2006) permit a role of the PMC in value processing and a causal role in choice behavior has been shown by microstimulation of this region in monkeys that leads to behavioral adaptation (Hayden et al., 2008). Additionally, the PMC has been suggested as part of a network tracking evidence for future adaptations to pending options (Boorman et al., 2011) in humans. Importantly, our results presented here differ from these previous findings, since we describe how the same stimulus value representation is updated by different signals depending only on whether feedback was

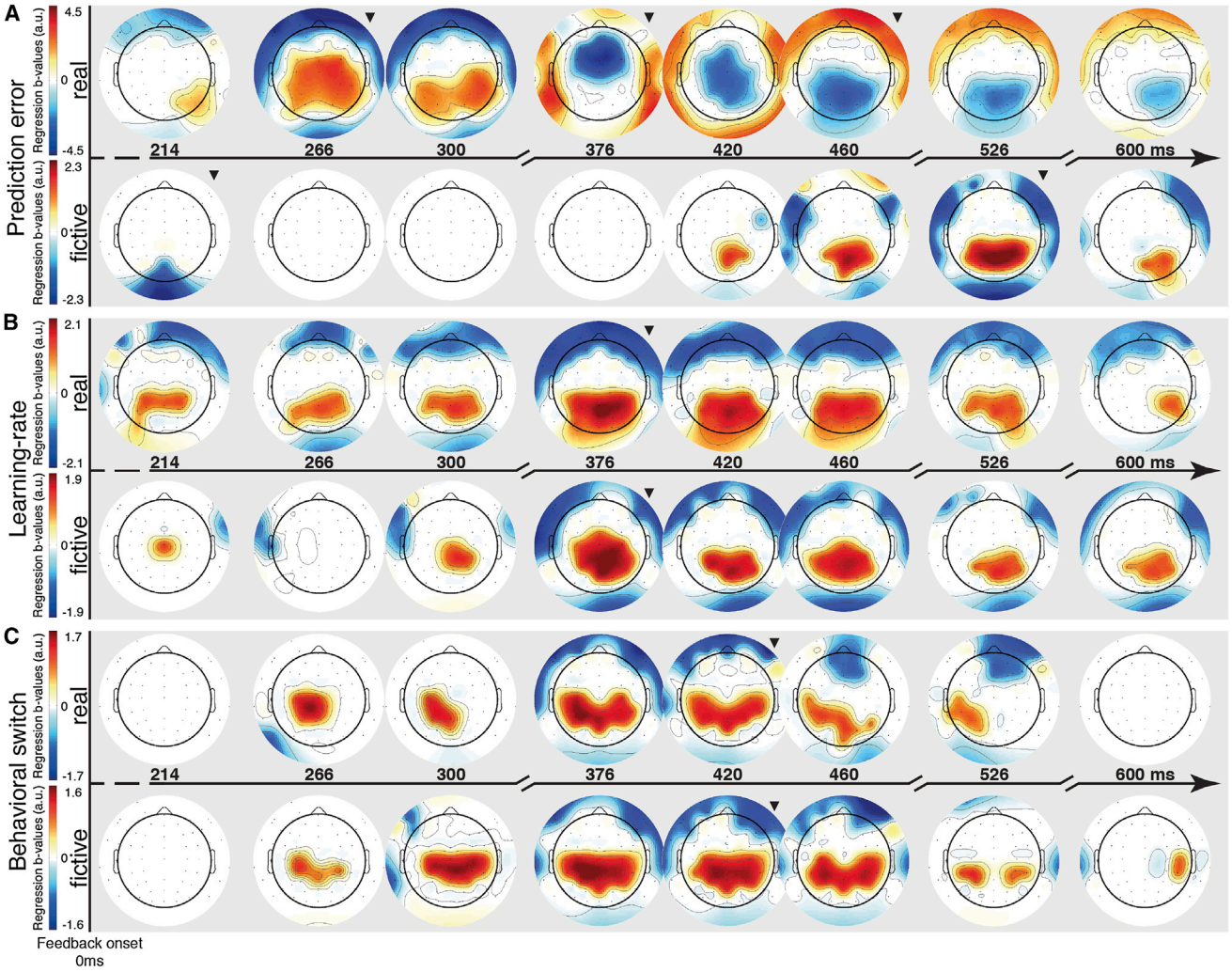

Figure 2. Multiple Robust Regression Analysis of Feedback-Locked Epochs

(A) Regression b values for feedback-locked results for the predictor PE plotted separately for real (top panels) and fictive feedback. Occipital electrodes showed the first significant effect only in the fictive condition (peak Oz 214 ms, $t_{30} = -5.78$, $p < 10^{-5}$). A frontocentral-positive (peak at FCz 266 ms, $t_{30} = 5.17$, $p < 10^{-4}$) and -negative (peak at FCz 382 ms, $t_{30} = -8.70$, $p < 10^{-8}$) effect were present only in the real feedback condition. Both conditions showed later parietal covariations that were opposed in sign (real peak at Pz 460 ms, $t_{30} = -7.86$, $p < 10^{-7}$ and fictive peak at Pz 526 ms, $t_{30} = 6.02$, $p < 10^{-5}$).

(B) In both feedback conditions, the learning rate showed a positive covariation with a centroparietal scalp distribution that peaked in a midlatency time window around 376 ms (at Pz for chosen $t_{30} = 8.38$, $p < 10^{-8}$ and for avoided $t_{30} = 7.41$, $p < 10^{-7}$).

(C) The behavioral switch regressor yielded a peak at parietooccipital electrodes (at Pz for chosen at 408 ms, $t_{30} = 7.12$, $p < 10^{-6}$; for avoided at 420 ms $t_{30} = 5.67$, $p < 10^{-5}$). Inverted triangles mark plots closest to peak amplitudes and nonsignificant ($p > 0.00069$) results are masked in white. See also Movie S1 for the whole time course of the effects and Figure S2 for source localizations.

fictive or real. We suggest that this signal might reflect a process that converts fictive outcomes to subjective value signals (Gold and Shadlen, 2007), effectively facilitating counterfactual learning that can more easily guide subsequent decisions.

This fictive PE effect cannot be interpreted as a surprise signal (Ferdinand et al., 2012), as it was unaffected when outcome and surprise, measured as the absolute PE value, were included into the same regression model (Figures S3E and S3F). Additionally, the effect cannot be interpreted as a consequence of repetition suppression (Summerfield et al., 2008), as it would then be expected to also occur following real feedback. In order to further disentangle contributing factors of the different PE correlates, we decomposed the PE into its components—the outcome and the expected value—and submitted both to the same multiple regression analysis. This revealed that the FRN in fact codes a PE signal, as both outcome and expected value showed significant effects with opposite signs indicating that the error signal increased when an unfavorable outcome was less expected (Figures S3A–S3D), thereby mimicking the response of dopaminergic neurons (Schultz et al., 1997). In contrast to this, the early fictive effect did not show significant influences by the expected value and thus may mainly code whether or not outcomes were

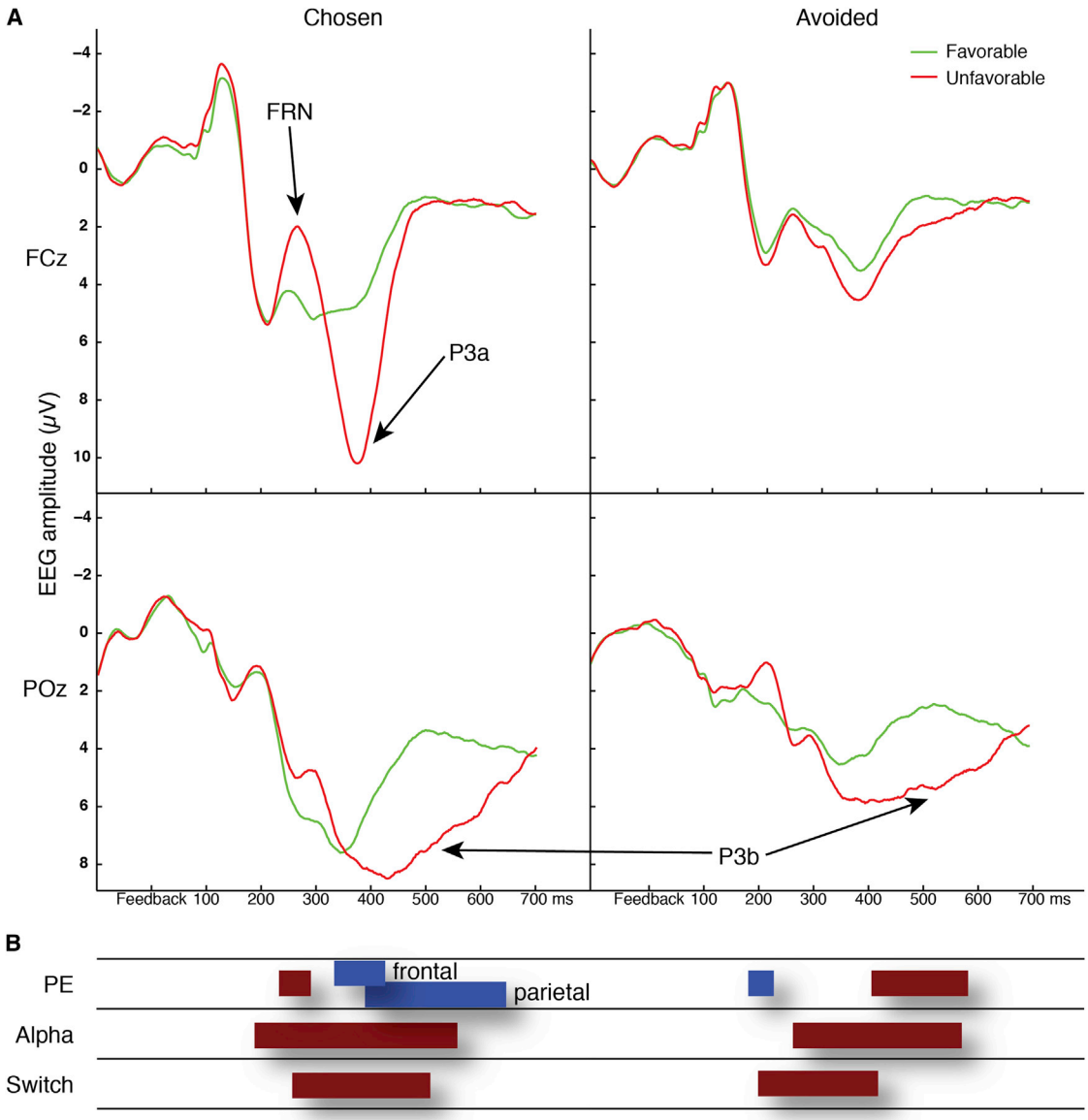

Figure 3. Comparison of Event-Related Potentials and Regression Results

(A) Grand average feedback-locked event-related potentials (ERPs) waveforms. Favorable (green) and unfavorable (red) outcomes are plotted separately for chosen and avoided feedbacks at electrode FCz and POz. A clear FRN component can be seen for real and fictive feedback at FCz but is not modulated by the valence of avoided feedback.

(B) Time windows of significant effects for regressors used in the regression analysis. Positive covariations are depicted in blocks in red and negative covariations in blue color. During the time window of the occipital early PE effect following fictive feedback, a negative deflection is present in the ERP waveform for fictive unfavorable feedback. Learning rate (alpha) and switch effects span over longer time windows that do not simply represent one single ERP component. See Figure S3 for a decomposition of the PE effects into contributing factors.

favorable. Neither P3a nor later components showed a pattern that satisfies criteria for an axiomatic PE signal (Caplin and Dean, 2008), which is in line with other studies that found the FRN to be the only cortical PE correlate in accordance with axioms of reward PE models (Talmi et al., 2012). Thus, the data suggest that different cortical areas covary with PEs at different times between 190–400 ms after feedback depending on whether an outcome is directly experienced or fictive. Furthermore, the very early occipital PE correlate is mainly driven by the favorability of the outcome itself and not the expected value, suggesting a binary evaluation taking place here that may later on be converted into more fine-scaled value updating.

Common Final Pathway

As feedback processing continues, the different streams appear to converge on a common late parietal PE correlate that coincides with the P3b ERP component (Polich, 2007). This PE covariation was evident in both conditions with reversed

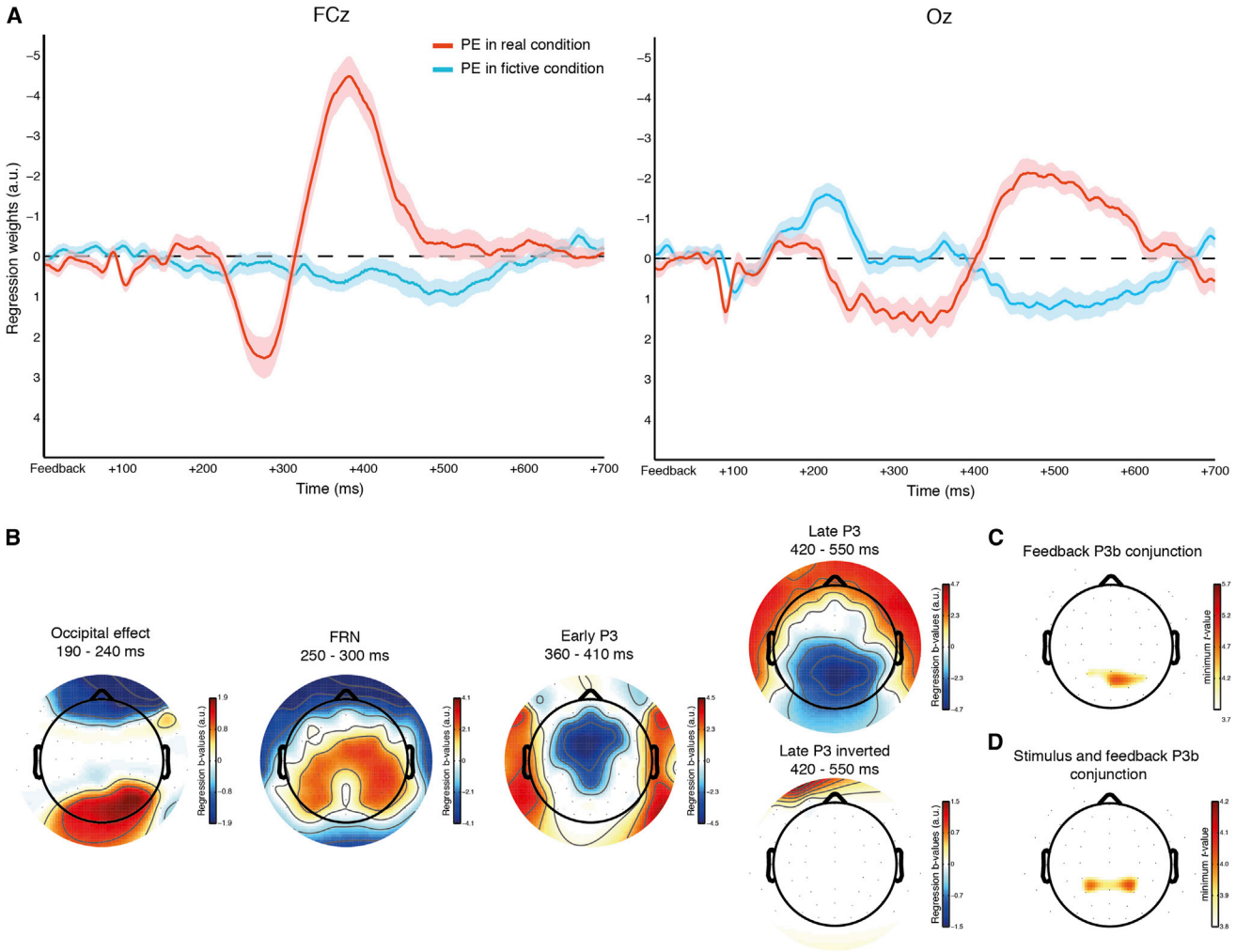

Figure 4. Time Courses of Regression Weights, Difference, and Conjunction Maps

(A) Time course of regression weights of the PE effects comparing processing of real (red) and fictive (blue) feedback, shown at electrodes of maximal effects of early PE correlates (FCz and Oz). Thick lines, mean regression weights; shadows indicate the SEM. See Figures S4D–S4F for an across-subjects correlation of regression weights with task performance.

(B) Difference-topography plots (fictive–real) in the time windows of the respective effects. Regression weights were collapsed over the effects' duration and average weights are plotted, while nonsignificant electrodes are masked in white ($p > 0.00069$). For the late parietal effect, because it occurred in both conditions, we also compared whether differences exist when the fictive condition was inverted (by multiplication with -1). This is based on the assumption that counterfactual thinking was employed following fictive outcomes, converting them to favorable or unfavorable events. Note that when inverted, both conditions did not differ significantly in the late P3b time window.

(C) Temporospatial conjunction map for regressors that showed effects in the late P3b time window in both real and fictive feedback conditions (learning rate, PE, and behavioral switch). Plotted are minimum *t*-statistics of significant coactivation of all regressors (Nichols et al., 2005) collapsed over time. Midline electrodes Pz and POz were significantly activated by all regressors in both conditions (see Figure S4 for details).

(D) Conjunction map for stimulus-locked SDC and feedback-locked PE and learning rate effects in both conditions between 370 and 650 ms. Parietal electrodes were coactivated by feedback- and stimulus-locked P3b effects for SDC in the decision making phase and by parameters critical for value updating in the feedback evaluation phase of the task.

polarities that were negative for real and positive for fictive feedback (significant from 392–650 ms for real and 414–590 ms for fictive feedback, Figures 2B and 3B). Notably, this polarity reversal results in the fact that unfavorable outcomes associated with negative PEs in real and positive PEs in fictive conditions always lead to positive-going deviations of parietal EEG activity. Thus, in order to compare the magnitude of the PE covariations, we multiplied the fictive feedback condition by -1 to account for the PE sign reversal in relation to the outcome's subjective valence before contrasting both conditions (Figure 4B). This is a logical consequence of the assumption that fictive feedback in which unfavorable outcomes are associated with positive PE signs engage counterfactual thinking. Contrasts did not show differences between conditions in this late time window, which indicates that real and fictive outcomes have similar effects on P3b modulations, although absolute P3b amplitudes are

reduced following fictive feedback (Figure 3). This effect might reflect the updating of stimulus-response mappings or, similarly, of a stimulus' expected value. Interestingly, an early theory of the P3b suggested that it covaries with deviations from an adaptation level (Ullsperger and Gille, 1988), a concept highly reminiscent of PEs, suggesting that the higher P3b is, the stronger the necessary deviation from default behavior. In line with this, the P3b amplitude was increased before a behavioral switch in a reversal-learning task (Chase et al., 2011). The P3b has been shown to correlate well with surprise (Mars et al., 2008), but surprise alone is insufficient to explain the late EEG modulation and behavioral switching in the present study: even when surprise was included as a separate regressor, P3b still displayed significant covariation in both conditions with the outcome itself (Figure S3). We thus suggest that the late parietal P3b effects modulated by PE represent a common pathway for adaptation based on the information extracted from the feedback. This view is strongly supported by the finding that an additional behavioral switch regressor (coding shift/stay behavior on next encounters with the same stimulus, which happened on average on the third following trial) covaries positively with midlatency and late parietal EEG amplitudes (Figure 2C), thus remarkably overlapping with the late PE effect in the temporal and spatial domain. Given previous findings that higher P3b amplitudes are associated with improved memory encoding (Fabiani et al., 1990; Paller et al., 1987), it is conceivable that the parietal EEG effects in the P3b time range reflect update and storage of the stimulus value. Intriguingly, the PE correlate appears longer lasting than the switch effect, suggesting that the late portions of the P3b may play further roles in addition to encoding the new stimulus value, speculatively autonomic responses and awareness (Wessel et al., 2011). In sum, the parietal P3b-like cortical activity seems to set the stage for future decisions.

As seen in the behavioral data and supported by the reinforcement learning model, with increasing learning success and thus increasing certainty of reward likelihood, the impact of feedback on value representations and overt behavioral adaptation decreases exponentially toward an asymptote. This is reflected in a decreasing learning rate α_t, which, regressed against EEG activity, yielded comparable sustained positive centroparietal effects (significant at Pz from 192–562 ms for real and from 272–580 ms for fictive feedback; Figure 2B). The maximal learning rate effect fell in between the early and the late PE effects in both conditions, thereby modulating the baseline of the FRN and the P3a and P3b amplitudes: the higher the learning rate the more positive the EEG signal. As the learning rate decreases, the EEG amplitude decreases as well. We suggest that this effect indeed represents the weighting of the outcome in both conditions, causing less value updating and behavioral adaptation in later trials within each block. This point is further corroborated by the observation that those subjects whose EEG signals more closely matched the reinforcement learning models' predictions made fewer bad decisions (Figures S4D–S4F). Our finding that the learning rate determines the baseline activity on which PE effects are modulated fits with fMRI results demonstrating that PE coding in pMFC is modulated by individual learning rates (Behrens et al., 2007; Jocham et al., 2009). Furthermore, it has been shown that functional connectivity of feedback processing brain areas is reduced in late phases of stable learning experiments (Klein et al., 2007).

Processing of Stimulus Information

When a stimulus value has been learned based on feedback, it needs to be retrieved and used to guide choice at the next encounter of the same stimulus. To investigate these processes, we submitted stimulus-locked EEG epochs to a multiple robust regression analysis. The signed Q_t regressor—reflecting the individual's single-trial stimulus value estimates—showed a significant positive covariation at frontal electrodes 250–268 ms after stimulus onset with peak values at electrode AFz (Figure 5). Thus, stimuli with higher subjective values were associated with more positive EEG activity. Value-related activity has consistently been reported to correlate with activity of the vmPFC (Jocham et al., 2012; Knutson et al., 2005; Plassmann et al., 2010; Wunderlich et al., 2010). The anterior distribution of this frontal value effect fits with an origin in vmPFC and its timing is supported by a recent study reporting vmPFC magnetoencephalic correlates of overall value when different stimuli were presented simultaneously (Hunt et al., 2012) and single-neuron activity in dlPFC and OFC in monkeys (Hayden et al., 2009). The translation of this value representation into action is indirect as indicated by an inverse relationship between EEG amplitude and reaction time for choosing compared to avoiding a stimulus (Figure S5A). This EEG modulation reflects the intuitive observation that Q values deviating further from 0 are associated with easier and quicker decisions about which option to choose (Figure S1A). In other words, choice reaction time is driven rather by the certainty of the stimulus value than by the value representation and its early EEG correlate.

Following this early covariation with signed value, a prominent effect of subjective decision certainty (SDC) about which response to give was seen. Values for SDC were derived from the likelihood of the computational model to select one response over the other and rectified in order to range from maximal uncertainty (0) to absolute preference of one option (1) (see Experimental Procedures for details). SDC demonstrated clear positive covariance with EEG activity in a centroparietal scalp distribution, peaking at around 520 ms following stimulus onset (significant from 456–744 ms, Figure 5), which is close to median response time (539 ms). Therefore, response certainty was reflected by more positive single-trial parietal EEG activity at a much later time point than the frontal value effects. The timing of the observed covariation fits well to the latency of the stimulus-related P3b ERP component. This pattern of increased P3b with response certainty rules out an explanation of novelty or surprise, as newly occurring stimuli always lead to SDC values of zero. Note that since RTs were included as a separate regressor in the multiple regression, neither the SDC nor the Q value effect can be explained by an earlier onset of preparatory motor activity (Figures S5A and S5C).

Predicting Future Adaptations

Our analysis enabled us to study the entire time course of cortical processes underlying decision making, outcome evaluation, and learning (i.e., updating) value representations. Upon stimulus presentation, retrieval of learnt values activates cortical value

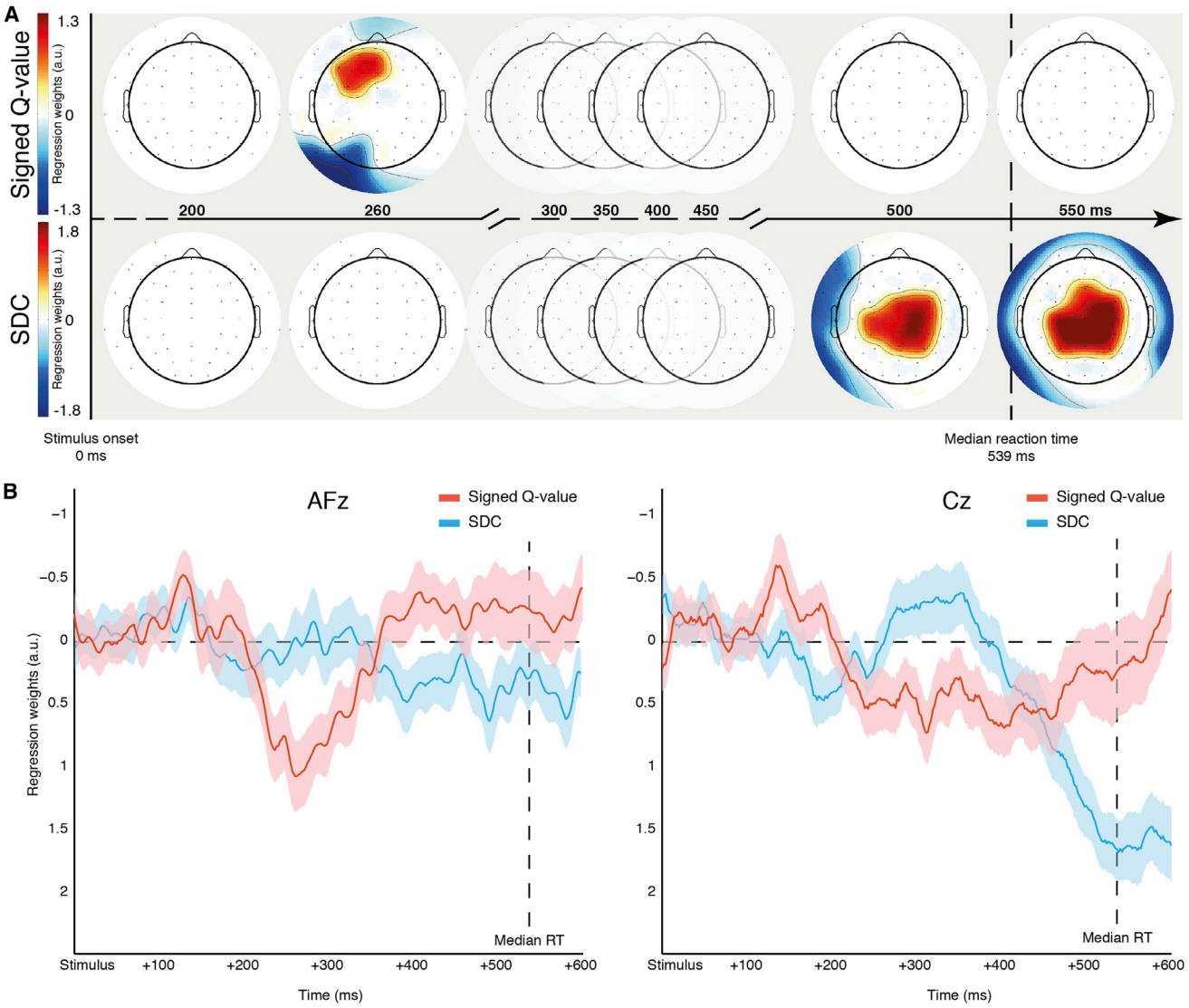

Figure 5. Stimulus-Locked Regression Results

(A) Top row: signed Q value showed a significant positive covariation at frontal electrodes (peak at Fz at 264 ms, $t_{30} = 4.06$, $p < 0.0005$, significant from 250 to 268 ms), lateralized to the left hemiscalp (AF3, F3/5, FC3). The scalp signal here was associated with reaction times in a condition-dependent manner (Figure S5B). Bottom row: results for the subjective decision certainty (SDC) regressor showed a positive mediocentral effect (peak at Cz at 520 ms, $t_{30} = 7.71$, $p < 10^{-7}$, significant from 456 to 744 ms). As SDC is higher when subjects are more certain about the response to give (for both good and bad stimuli), these results imply increased stimulus P3b amplitudes after subjects established reasonable certainty about expected values and thus the optimal response. This effect is independent of whether or not the stimulus was estimated to be good or bad as in this time window no effect of the SQV was observed.

(B) Corresponding time courses of regression weights. SQV and SDC at the two electrodes of their respective peak values (AFz and Cz) are shown. Thick lines, mean regression weights; shadows indicate the SEM. The vertical dashed line represents the group median reaction time (539 ± 16 ms) in both plots. Nonsignificant ($p > 0.00044$) time points in the topography plots are masked in white. See also Figure S5C for response-locked results.

representations reflected in early midfrontal EEG activity. Decision certainty is reflected in P3b-like parietal EEG activity around response latency, and mapping of the selected action to the motor response is reflected in lateralized activity from (pre)motor cortices (Figure S5C). After feedback, initially outcomes are processed separately depending on whether their consequences are real or fictive, presumably in order to convert feedback information into a common value currency allowing for efficient learning of stimulus values. Then the information about necessary value updates converges on common parietal P3b-like activity modulated by whether the action was successful or not. Given the probabilistic nature of the instrumental learning task, several parameters need to be used to weight the impact of single-trial outcomes. Over the course of multiple trials, learning rate indicates the learning success and downweights the single-feedback information at later learning stages. Moreover, when a choice is made with high certainty, perseveration of this behavior is favorable. This means that already at the

time of the response (and thus before feedback), high certainty might be used to strengthen the current value representation, thereby shielding it from potentially misleading feedback. Interestingly, the stimulus- and feedback-locked late parietal P3b-like activity is consistent with the notion of certainty- and learning-rate-weighted value strengthening and updates at different time points: high response certainty, which should be associated with re-encoding (strengthening) of the stimulus value to assure perseveration, is associated with high stimulus-locked P3b amplitudes. In contrast, after feedback, high learning rates and unfavorable outcomes commonly give rise to high feedback-locked P3b amplitudes, presumably reflecting value updating and storage, thereby increasing the likelihood to change future choice behavior. To put it briefly, lower stimulus-related P3b and higher feedback-related P3b amplitudes should be associated with an increased likelihood to switch choice on the next encounter with the same stimulus.

This notion that feedback- and stimulus-related P3b amplitudes are inversely related to switch behavior was tested at electrode Pz, which was identified via a conjunction analysis of all relevant stimulus- and feedback-locked effects in the P3b time window (Figure 4D). A discrimination threshold was iteratively estimated in one half of randomly chosen trials that was then used to predict switching in the second half of trials. This very simple algorithm predicted switches significantly above chance level, namely with average accuracy of 56.75% ± 1.18% (t_{30} = 5.67, $p < 10^{-5}$) following real outcomes and 55.86% ± 0.72% (t_{30} = 8.26, $p < 10^{-8}$) following fictive outcomes. When this algorithm was applied to the stimulus-related P3b, switches were predicted correctly with average accuracy of 53.17% ± 0.78% (t_{30} = 4.04, p = 0.0003) before choosing and 53.36% ± 0.77% (t_{30} = 4.34, p = 0.0001) before avoiding the stimulus. Note that the purpose of this analysis was not to predict future behavior as accurately as possible but to demonstrate that the whole-brain regression reliably identified electrodes and time windows of importance for studying learning and decision making and that switches still refer to the next time the stimulus is shown again. Importantly, it was indeed the case that switches were predicted by increased feedback-related but decreased stimulus-related P3b amplitudes (see Experimental Procedures for details). This result demonstrates that simple attentional effects cannot account for the P3b effects: a global decrease of attention should lower stimulus- and feedback-related P3b amplitudes (Polich, 2007) and adaptive switches in parallel, which is inconsistent with our findings. To compare the importance of both factors in predicting future adaptations, we used logistic regression on the switch behavior to determine the contributions of stimulus and feedback P3b. When let to compete for variance, feedback P3b was the better indicator of behavioral adaptation (p = 0.035 for chosen and p = 0.028 for avoided stimuli, two-sided t test of standardized regression weights), but both feedback and stimulus P3b had a significant effect (all $p < 0.01$). As is intuitively plausible, the actual feedback is more closely related to adaptation but already before feedback is presented, predictions about behavioral adaptation based solely on stimulus values are possible. Thus, with the mere knowledge of a short interval of raw stimulus- or feedback-related EEG at Pz and current behavior, predictions of future behavior can be made.

This strengthens the interpretation of feedback P3b representing value updating, as P3b in both stages of decision making alludes to value coding and behavioral adaptation. It is tempting to assume that both processes are related and that, in case of high certainty, already before feedback is given the stimulus value is encoded. Although similarity in both processes is suggested by the conjunction analysis, these EEG results have to be interpreted cautiously as different generators may give rise to similar scalp topographies. The reversal of the relationship between P3b amplitudes and switch behavior, however, hints to a more specific mechanism than a mere reduction of attention or simple surprise. It therefore seems to be the case that PE correlates, processed in different cortical areas for real and fictive outcomes, modified by a weighting process, serve as the basis for, and precede the timing of, future decisions. After being exposed to an updated stimulus again, P3b covaries with the security of the selected action, possibly preventing switching away from a learned stimulus.

Being able to adapt behavior based on purely fictive events through counterfactual thinking may be a human ability that allows learning from abstract information in the absence of any actor. Our results demonstrate through the whole time course of decision making, from value retrieval following stimulus presentation and its translation into action selection until the updating of these values following feedback, how real and fictive events can be utilized to enable adaptive behavior. Localization and timing of these fictive error signals suggest a distinct function that may have evolved by recruiting different cortical mechanisms than experiencing or observing real outcomes caused by an actor. The adaptation itself, however, seems to be based on a more general mechanism that can be employed by experienced and fictive outcomes.

EXPERIMENTAL PROCEDURES

Participants

Thirty-one healthy subjects (21 female, mean age: 23.81 ± 0.61) participated in a pharmacological study and each provided written informed consent. We report here on data from the placebo session. The study was approved by the ethics committee of the Medical Faculty of the University of Cologne (Cologne, Germany).

Task Description

Subjects had to learn the associated reward probabilities of different stimuli in order to maximize their financial earnings in a probabilistic choice task. At each trial, subjects were presented with one stimulus where they had two options: they could either choose the stimulus and risk winning or losing €0.10 or avoid the stimulus and observe the outcome without financial consequences. The fictive feedback provided information about what would have happened if they had chosen that stimulus (fictive outcome). Subjects were informed that they would receive the money won in the task at the end of the session as a bonus to their expense allowance. The task was presented using Presentation 10.3 (Neurobehavioral Systems).

The experiment consisted of four blocks with a random series of three different stimuli, totaling 12 different stimuli over the time of the experiment. Four stimuli associated with high chances of reward (*good* stimuli, two with 80% and two with 70% win rate), four stimuli associated with low chances of reward (*bad* stimuli, with 20% and 30% win rate), and four stimuli with a random chance of winning (*neutral* stimuli, 50% win rate) were presented 50 times each and then replaced. Win rates and symbol sequences were pseudorandomized. There were no pauses during the experiment, and trials in

which subjects failed to respond within the given deadline were discarded from analysis. In the last block of the experiment, until each stimulus had been shown 50 times, additional new filler stimuli were shown but not included in the analyses so that every subject concluded exactly 600 valid trials.

Each trial began with a central fixation cross that was shown for a random period between 300 and 700 ms accompanied by the two response options: *choose* (indicated by a green tick mark) and *avoid* (indicated by a red no parking sign, Figure 1A). The response options remained in place until feedback was shown and their sides were counterbalanced across subjects. After the fixation cross, one central stimulus consisting of drawn animal pictures in white on a black background was presented until the subject responded or 1,700 ms had elapsed. If subjects failed to respond in time, a message appeared asking them to respond faster. Subjects' choices were confirmed by a white rectangle surrounding the chosen option for 350 ms. Immediately thereafter, the outcome was presented for 750 ms depending on the subjects' choice. If subjects bet money, they received either a green smiling face and a reward of €0.10 or a red frowning face and a loss of €0.10. When subjects did not bet on a symbol, they received the same feedback but with a slightly paler color and the money that could have been received was crossed out to indicate that the feedback was fictive and had no monetary effect. Stimuli were kept as similar as possible between conditions to avoid introducing effects of stimulus salience. On average, subjects gained €6.36 ± €0.51 (range €0.50–€9.50) over the course of the experiment.

EEG Data Acquisition and Analysis

Scalp voltages were recorded with 60 Ag/AgCl sintered electrodes from participants seated in a dimly lit electromagnetically and acoustically shielded chamber. Electrodes were mounted in an elastic cap (Easycap) in the extended 10-20 system with impedances kept below 5 kΩ. The ground electrode was positioned at F2 and data were online referenced to electrode CPz. Eye movements were captured by electrodes positioned at the left and right outer canthus and above and below the left eye, respectively. EEG data were registered continuously at 500 Hz sampling frequency with BrainAmp MR plus amplifiers (Brain Products). Data were then offline analyzed using EEGLAB 7.2 (Delorme and Makeig, 2004) and custom routines in MATLAB 7.8 (MathWorks). After filtering the signal from 0.5 to 52 Hz and rereferencing to common average reference, epochs spanning from −1.5 s before to 1.5 s after feedback and −1 s before to 1 s after stimulus onset were generated. Epochs containing deviations greater than 5 SD of the mean probability distribution on any single channel or the whole montage were automatically rejected. Epoched data were then submitted to temporal infomax independent component analysis (ICA) integrated in EEGLAB and manually corrected for artifacts such as eye blinks. Hereafter, data were re-epoched to extract response-locked data with epochs spanning from −500 ms before until 100 ms after the response. The average EEG activity spanning from −250 to −50 ms before stimulus and feedback presentation and −500 to −400 ms before response onset was used as baseline and subtracted from each channel individually (see Supplemental Experimental Procedures for additional results of the stimulus- and response-locked data).

Computational Model

We used a reinforcement Q-learning algorithm to model each subject's sequence of choices (Sutton and Barto, 1998), which has been successfully adopted in reinforcement-learning paradigms (e.g., Jocham et al., 2009). For each stimulus and trial t, the model estimated the expected stimulus value Q_t based on that stimulus' previous reward and choice history. Q values represent the expected reward (positive values) or punishment (negative values) and are updated according to the following rule:

$$Q_{t+1} = \begin{cases} Q_t + \alpha_{c,\,t}\delta_t \text{ if chosen} \\ Q_t + \alpha_{a,\,t}\delta_t \text{ if avoided} \end{cases}. \quad (1)$$

δ_t represents the PE of the given trial, calculated as the difference between Q value and reward magnitude (R_t):

$$\delta_t = \frac{R_t - Q_t}{2}. \quad (2)$$

To update the Q value in Equation (1), we scaled the amplitude of δ_t by exponentially decreasing learning rates $\alpha_{c,t}$ and $\alpha_{a,t}$, respectively, depending on whether the subject had chosen or avoided the stimulus. This allowed assessment of differences in learning rates and behavioral flexibility on both conditions separately. The exponential decay was calculated by two half-life time parameters ($Hl_{c/a}$) depending on the subject's choice:

$$\alpha_{c,\,t} = \frac{\alpha_{c,\,1}}{2^{\left(\frac{t-1}{Hl_c}\right)}} \text{ and } \alpha_{a,\,t} = \frac{\alpha_{a,\,1}}{2^{\left(\frac{t-1}{Hl_a}\right)}}. \quad (3)$$

$\alpha_{c,1}$ and $\alpha_{a,1}$ denote the two free parameters representing the initial learning rate in both conditions. A lower limit for $\alpha_{c,t}$ and $\alpha_{a,t}$ was set to 0.01, under which learning rates could not decrease. Note that our model additionally contained a constant learning rate ($Hl_{c/a} = \infty$) as part of the range of parameters in the fitted parameter set to account for the possibility of a time invariant learning rate.

The likelihood of the model to choose or avoid a given stimulus was calculated by the softmax rule of the associated Q value (Figure 1B):

$$P_{c,\,t} = \frac{1}{1 + exp(-Q_t\beta)} \text{ and } P_{a,\,t} = 1 - P_{c,\,t}. \quad (4)$$

The free sensitivity parameter β can be regarded as the inverted temperature (high values lead to predictable behavior and vice versa). For the first step, we determined parameter estimates for all five free parameters using a grid search minimizing $-LL$ over all trials T:

$$nLL = \sum_{t=1}^{T} log\, P(c_t|\theta). \quad (5)$$

$P(c_t|\ominus)$ denotes the models' probability to choose in the same way as the subject did in each trial given the parameter-set theta. To determine reasonable parameter combinations, we applied the following constraints: $\alpha_{c/a,1} \geq 0.01$ and ≤ 1, $Hl_{c/a} \geq 1$ and ≤ 100 but separately including ∞ and $\beta \geq 0.01$ and ≤ 25 and step sizes for β were logarithmized. The logarithmization reflects the assumption that the model is more strongly affected by differences at small β values. Second, the best-fitting parameter combination was then used as the starting point for a nonlinear optimization algorithm (fmincon, MATLAB optimization toolbox). Constraints for $\alpha_{c,1}$ and $\alpha_{a,1}$ were kept but no upper limits for β and $Hl_{c/a}$ set. To obtain single-trial estimates of δ_t, Q_t, and $\alpha_{c/a,t}$, we re-entered the MLE parameters $\alpha_{c/a,t}$, $Hl_{c/a}$, and β into the reinforcement-learning algorithm.

Model Fit and Parameters

The parameter combinations that led to the best fit were not significantly different between both conditions (Table 1). Best fits were obtained for slightly higher average initial learning rates in condition choose ($\alpha_{c,1} = 0.48 \pm 0.07$) than in avoid ($\alpha_{a,1} = 0.42 \pm 0.07$), which decreased slightly more rapidly ($Hl_c = 9.78 \pm 2.60$ and $Hl_a = 13.47 \pm 3.30$). For one subject, the best fit was obtained with a constant learning rate (defined as a half-life time >100 trials, which equals less than ~30% decrease per block) in condition choose and for four subjects in condition avoid. On average, learning rates decreased to 3% of their initial values in condition choose and to 8% in condition avoid, providing strong support for the assumption that the impact of PEs is reduced over time. To compare both learning rates between conditions, we conducted a repeated-measures ANOVA with factors α_t *(50)* and *condition (2)* that showed no significant main effect of *condition* on the decaying learning rate (*condition* $F_{1,30} = 0.26$, $p = 0.613$) and no interaction (*condition x* α_t $F_{1.8,54} = 0.553$, $p = 0.561$).

Although we fit different sets of model parameters for both conditions (real and fictive), we did not account for possible differences in learning caused by the different reward contingencies. It is likely that this would influence the results for parameter MLE, especially for the decaying learning rate. Notably, we did not observe a significant feedback-locked effect for the decaying learning rate when analysis was restricted to neutral stimuli alone, indicating that here no downweighting of the PEs in later trials occurred (see Supplemental Experimental Procedures). However, we feel that fitting parameters separately, even for different reward contingencies, would lead to overfitting and expand parameter space to unmanageable dimensions.

Multiple Single-Trial Robust Regression

To account for differences in the sensitivity parameter, *Z* scored results of the reinforcement-learning model were used to build a general linear model (GLM) and regress single-trial EEG activity at each electrode and time point against model predictions and behavioral parameters. Robust regression that down-weights outliers by performing an iteratively reweighted least square method (O'Leary, 1990) was employed to determine parameters in the following linear equation: $Y = intercept + b_1Reg_1 + b_2Reg_2 \ldots + error$.

Similar approaches have been successfully applied to EEG time- (Rousselet et al., 2008) and frequency-domain (Cohen and Cavanagh, 2011) data and allow the simultaneous investigation of multiple independent variables while preserving the high temporal resolution of the EEG. This mass univariate approach leads to individual *b* values for each electrode and time point for every subject. To ensure comparability between predictors within and between subjects and to penalize the model in case of multicollinearity of predictors, *b* values were standardized by their SDs before averaging across subjects.

The stimulus-locked GLM included variable learning-rate (α_t), signed Q value (sQ_t), and subjective decision certainty (*SDC*) plus the reaction time (*RT*) as a regressor of no interest. SDC was calculated from the models' softmax likelihood by equalizing $P_{c/a}$ for choosing and avoiding using the following equation: $SDC = \text{abs}(P_{c/a} - 0.5) \times 2$. The result ranged from 0 (maximal insecurity) to 1 (absolute preference of one option). Feedback-locked data were analyzed separately for the categorical conditions *fictive* and *real*. Predictors included the PE (δ_t), variable learning rate (α_t), and a dichotomous regressor indicating a switch of response (coded as 1) or a stay (coded as 0) on the next trial that the same stimulus was shown again.

Standardized *b* values can be assumed to be Gaussian due to the central limit theorem and thus could be tested via two-tailed one-sample t tests, which were done separately at each data point in a whole-brain approach across subjects. Resulting p values were corrected for multiple comparisons using false discovery rate (FDR) following the method suggested by Benjamini and Yekutieli (2001), which has been shown to provide solid control of the family-wise error rate (FWER) in EEG data (Groppe et al., 2011). However, as FDR in itself does not provide strong (local) control of the FWER, it was applied to all concatenated *b* value data sets per model. This ensured that all corrections were done with the same threshold value for each regressor in the models. H_0 was rejected for all $p < 0.00070$ in the feedback and $p < 0.00045$ in the stimulus-locked model. Nonsignificant data points are masked in white in the topography plots and Movie S1. Both conditions in the feedback-locked epochs were contrasted via paired two-tailed t tests thresholded at the same level as noted above.

Direct Contrasts between Real and Fictive Feedback Evaluation

We compared both real and fictive feedback processing directly via paired two-sided t tests of the regression *b* values, thresholded at the same level determined by FDR. This revealed that feedback processing indeed differed significantly for all PE effects. The late parietal effect did not differ significantly when it was inverted for fictive feedbacks, assuming that counterfactual thinking was employed (by multiplication with −1) before contrasting. Contrasts for alpha and switch regressors did not reveal significant differences between both conditions.

Prediction of Choice Switches Based on Artifact-free Raw EEG

Artifact-free raw EEG was averaged from 370 to 430 ms at electrode (Pz) that showed the biggest overlap between effects of the switch, PE, and learning rate predictors in the regression analysis (Figures 4C, 4D, and S4) and SDC effects locked to stimulus onset. As we observed a positive covariation in the regression analysis for switching behavior, we hypothesized that higher EEG amplitudes should be associated with a higher likelihood to switch. Additionally, because the absolute EEG amplitudes differed between both conditions (Figure 3), the analyses for real and fictive feedback were performed separately. For each subject, equally sized samples were randomly drawn from both conditions and split into two halves. One half was used to determine a discrimination threshold calculated as the simple arithmetic mean between the distributions of amplitudes for switches and stays. The predictions of this threshold were then tested in the other half of trials. One hundred iterations were performed, and the results were averaged and tested against chance (i.e., 50% correct predictions) on group level using two-tailed one-sample t tests.

Average amplitudes in the defined time window before switches were 6.29 ± 0.71 μV and 4.03 ± 0.62 μV before stays following chosen and 4.00 ± 0.61 μV before switches and 1.78 ± 0.53 μV before stays following fictive feedback. Predictions were equally valid in both conditions: the simple discrimination algorithm predicted switches correctly on an average of 56.75% ± 1.18% (t_{30} = 5.67, $p < 10^{-5}$) for real feedback and 55.86% ± 0.72% (t_{30} = 8.26, $p < 10^{-8}$) for fictive feedback (t_{30} = 0.68, p = 0.49 for difference in accuracy between conditions). In the real feedback condition out of the 31 participants, 27 had a prediction chance >50% and 17 >55% (maximum 70.88%). In the fictive feedback condition, 29 had a prediction chance >50% and 17 >55% (maximum 65.76%). Results remained significant when only neutral, good, or bad stimuli were analyzed (always $p < 0.01$).

The same analysis was performed for stimulus-locked data, again separately for upcoming choose or avoid decisions to keep the results comparable with the feedback locked analysis. Average P3b amplitudes (from 520 to 580 ms, measured at Pz) before choosing were 4.81 ± 0.41 μV and 5.77 ± 0.48 μV for stimuli that lead to switches and stays, respectively. Before avoiding, average amplitudes were 4.78 ± 0.46 μV before switches and 5.99 ± 0.46 μV for stimuli that lead to switches and stays, respectively. Switches were predicted correctly on an average of 53.17% ± 0.78% (t_{30} = 4.04, p = 0.0003) for real feedback and 53.36% ± 0.77% (t_{30} = 4.34, p = 0.0001) for fictive feedback. Before choosing out of the 31 participants, 24 had a prediction chance >50% and 8 >55% (maximum 61.38%) and before avoiding 26 had a prediction chance >50% and 11 >55% (maximum 61.62%). Results remained significant when only good or bad stimuli were analyzed in both conditions (always $p < 0.01$) but not when only neutral stimuli were analyzed (both $p > 0.39$). The latter has to be expected as it is implausible that it would be possible to predict future adaptations for random outcomes if these affect switching. No differences were seen in the latency of the grand-average peak of the stimulus P3b amplitudes depending on high or low expected values or on following choices (ANOVA choice (choose/avoid) × value (high/low) p always > 0.1).

Logistic regression of switch behavior against stimulus and feedback P3b amplitudes was used to compare their respective predictive powers. Standardized *b* values for choices were 1.7 ± 0.26 (t_{30} = 6.55, $p < 10^{-6}$) for feedback and −0.66 ± 0.22 (t_{30} = −3.05, p = 0.005) for stimulus P3b amplitudes. For avoided trials, *b* values were 1.72 ± 0.20 (t_{30} = 8.59, $p < 10^{-8}$) for feedback and −0.79 ± 0.23 (t_{30} = −3.45, p = 0.0016) for stimulus P3b amplitudes. Note the sign reversal of regression weights for stimulus and feedback P3b in relation to switch behavior. Combining feedback- and stimulus-locked P3b amplitudes did not increase prediction accuracy for the logistic regression as measured by comparing summed −*LL* via likelihood-ratio tests between the model with only feedback P3b and the combined model (both $p > 0.59$).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2013.07.006.

ACKNOWLEDGMENTS

We thank Gerhard Jocham, Theo O.J. Gründler, and Tanja Endrass for fruitful discussions on the presented data and Sabrina Döring for support in data collection. This work was supported by grants of the German Ministry of Education and Research (BMBF, 01GW0722) and from the German Research Foundation (DFG, SFB 779 A 12) to M.U.

Accepted: July 11, 2013
Published: September 18, 2013

REFERENCES

Amiez, C., Joseph, J.P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. Cereb. Cortex *16*, 1040–1055.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat. *29*, 1165–1188.

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

Burke, C.J., Tobler, P.N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. Proc. Natl. Acad. Sci. USA *107*, 14431–14436.

Caplin, A., and Dean, M. (2008). Axiomatic methods, dopamine and reward prediction error. Curr. Opin. Neurobiol. *18*, 197–202.

Chase, H.W., Swainson, R., Durham, L., Benham, L., and Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. J. Cogn. Neurosci. *23*, 936–946.

Chiu, P.H., Lohrenz, T.M., and Montague, P.R. (2008). Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat. Neurosci. *11*, 514–520.

Cohen, M.X., and Cavanagh, J.F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. Front. Psychol. *2*, 30.

de Bruijn, E.R.A., de Lange, F.P., von Cramon, D.Y., and Ullsperger, M. (2009). When errors are rewarding. J. Neurosci. *29*, 12183–12186.

Delorme, A., and 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.

Fabiani, M., Karis, D., and Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. Electroencephalogr. Clin. Neurophysiol. *75*, 22–35.

Ferdinand, N.K., Mecklinger, A., Kray, J., and Gehring, W.J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. J. Neurosci. *32*, 12087–12092.

Gehring, W.J., and Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. Science *295*, 2279–2282.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Groppe, D.M., Urbach, T.P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology *48*, 1711–1725.

Gruendler, T.O.J., Ullsperger, M., and Huster, R.J. (2011). Event-related potential correlates of performance-monitoring in a lateralized time-estimation task. PLoS ONE *6*, e25591.

Hayden, B.Y., Nair, A.C., McCoy, A.N., and Platt, M.L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. Neuron *60*, 19–25.

Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2009). Fictive reward signals in the anterior cingulate cortex. Science *324*, 948–950.

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

Hunt, L.T., Kolling, N., Soltani, A., Woolrich, M.W., Rushworth, M.F.S., and Behrens, T.E.J. (2012). Mechanisms underlying cortical activity during value-guided choice. Nat. Neurosci. *15*, 470–476, S1–S3.

Jocham, G., Hunt, L.T., Near, J., and Behrens, T.E.J. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. Nat. Neurosci. *15*, 960–961.

Jocham, G., Neumann, J., Klein, T.A., Danielmeier, C., and Ullsperger, M. (2009). Adaptive coding of action values in the human rostral cingulate zone. J. Neurosci. *29*, 7489–7496.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*, 940–947.

Klein, T.A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D.Y., and Ullsperger, M. (2007). Genetically determined Differences in Learning from errors. Science *318*, 1642–1645.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. J. Neurosci. *25*, 4806–4812.

Leichnetz, G.R. (2001). Connections of the medial posterior parietal cortex (area 7m) in the monkey. Anat. Rec. *263*, 215–236.

Lohrenz, T., McCabe, K., Camerer, C.F., and Montague, P.R. (2007). Neural signature of fictive learning signals in a sequential investment task. Proc. Natl. Acad. Sci. USA *104*, 9493–9498.

Mars, R.B., Jbabdi, S., Sallet, J., O'Reilly, J.X., Croxson, P.L., Olivier, E., Noonan, M.P., Bergmann, C., Mitchell, A.S., Baxter, M.G., et al. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. J. Neurosci. *31*, 4087–4100.

Mars, R.B., Debener, S., Gladwin, T.E., Harrison, L.M., Haggard, P., Rothwell, J.C., and Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. J. Neurosci. *28*, 12539–12545.

Miltner, W.H.R., Braun, C.H., and Coles, M.G.H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. J. Cogn. Neurosci. *9*, 788–798.

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage *25*, 653–660.

Nieuwenhuis, S., Holroyd, C.B., Mol, N., and Coles, M.G.H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. Neurosci. Biobehav. Rev. *28*, 441–448.

O'Leary, D.P. (1990). Robust regression computation using iteratively reweighted least squares. SIAM J. Matrix Anal. Appl. *11*, 466–480.

Paller, K.A., Kutas, M., and Mayes, A.R. (1987). Neural correlates of encoding in an incidental learning paradigm. Electroencephalogr. Clin. Neurophysiol. *67*, 360–371.

Parvizi, J., Van Hoesen, G.W., Buckwalter, J., and Damasio, A. (2006). Neural connections of the posteromedial cortex in the macaque. Proc. Natl. Acad. Sci. USA *103*, 1563–1568.

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

Plassmann, H., O'Doherty, J.P., and Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. J. Neurosci. *30*, 10799–10808.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. *118*, 2128–2148.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443–447.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. *2*, 661–670.

Rousselet, G.A., Pernet, C.R., Bennett, P.J., and Sekuler, A.B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. BMC Neurosci. *9*, 98.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Subiaul, F., Cantlon, J.F., Holloway, R.L., and Terrace, H.S. (2004). Cognitive imitation in rhesus macaques. Science *305*, 407–410.

Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. *11*, 1004–1006.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: an Introduction. (Cambridge: MIT Press).

Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., and Dolan, R.J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. Neuroimage *59*, 635–645.

Ullsperger, M., and von Cramon, D.Y. (2003). Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. J. Neurosci. *23*, 4308–4314.

Ullsperger, P., and Gille, H.G. (1988). The late positive component of the ERP and adaptation-level theory. Biol. Psychol. *26*, 299–306.

van Schie, H.T., Mars, R.B., Coles, M.G.H., and Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. Nat. Neurosci. *7*, 549–554.

Walton, M.E., Devlin, J.T., and Rushworth, M.F.S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. Nat. Neurosci. *7*, 1259–1265.

Watkins, C.J.C.H., and Dayan, P. (1992). Q-Learning. Mach. Learn. *8*, 279–292.

Wessel, J.R., Danielmeier, C., and Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. J. Cogn. Neurosci. *23*, 3021–3036.

Wunderlich, K., Rangel, A., and O'Doherty, J.P. (2010). Economic choices can be made using only stimulus values. Proc. Natl. Acad. Sci. USA *107*, 15005–15010.

Yu, R., and Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. Neuroreport *17*, 1747–1751.