Neuropsychologia 66 (2015) 75–87

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Processing of action- but not stimulus-related prediction errors differs between active and observational feedback learning

Stefan Kobza [a,*], Christian Bellebaum [b]

[a] Institute of Cognitive Neuroscience, Department of Neuropsychology, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44780 Bochum, Germany
[b] Institute for Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form
19 September 2014
Accepted 27 October 2014
Available online 7 November 2014

Keywords:
Reward
Learning
Observation
Striatum
Prediction error
Cerebellum

ABSTRACT

Learning of stimulus–response–outcome associations is driven by outcome prediction errors (PEs). Previous studies have shown larger PE-dependent activity in the striatum for learning from own as compared to observed actions and the following outcomes despite comparable learning rates. We hypothesised that this finding relates primarily to a stronger integration of action and outcome information in active learners. Using functional magnetic resonance imaging, we investigated brain activations related to action-dependent PEs, reflecting the deviation between action values and obtained outcomes, and action-independent PEs, reflecting the deviation between subjective values of response-preceding cues and obtained outcomes. To this end, 16 active and 15 observational learners engaged in a probabilistic learning card-guessing paradigm. On each trial, active learners saw one out of five cues and pressed either a left or right response button to receive feedback (monetary win or loss). Each observational learner observed exactly those cues, responses and outcomes of one active learner. Learning performance was assessed in active test trials without feedback and did not differ between groups. For both types of PEs, activations were found in the globus pallidus, putamen, cerebellum, and insula in active learners. However, only for action-dependent PEs, activations in these structures and the anterior cingulate were increased in active relative to observational learners. Thus, PE-related activity in the reward system is not generally enhanced in active relative to observational learning but only for action-dependent PEs. For the cerebellum, additional activations were found across groups for cue-related uncertainty, thereby emphasising the cerebellum's role in stimulus–outcome learning.

1. Introduction

One of the most important principles of evolution is the selection of traits (of a species or an individual) which fit best in the current environmental conditions. Accordingly, organisms seek to select behaviour which fits best in the current situation, that is, behaviour which is followed by the most positive outcome. In order to maintain a high proportion of positive outcomes in an ever-changing environment, the ability to adapt behaviour is crucial. This adaptive process is reflected in an increase and decrease of the probability of behaviour followed by positive and negative outcomes, respectively, suggesting a preceding learning process: for example, the extent to which an outcome is worse than predicted may serve as a 'teaching signal' to select an alternative action in the future. Conversely, an outcome better than predicted may elicit a signal which reinforces repetition of the preceding action. Animal and human studies have revealed neural correlates of these so-called prediction errors (PEs): Dopamine (DA) neurons in the monkey fire at higher frequency following unexpected reward, whereas the firing rate drops below baseline when expected reward is omitted (Schultz, 1997, 1998a, 1998b; Schultz et al., 1997). A similar pattern was found via microelectrode recordings in Parkinson's Disease (PD) patients during deep brain surgery (Zaghloul et al., 2009). Further evidence in humans stems from functional magnetic resonance imaging (fMRI) studies showing outcome-related activations also in brain regions receiving projections from midbrain DA neurons (Haber and Fudge, 1997), most prominently the basal ganglia (BG; Delgado, 2007; Pagnoni et al., 2002) and the medial prefrontal cortex (mPFC) (O'Doherty et al., 2001; Rolls et al., 2008), particularly the anterior cingulate cortex (ACC; Holroyd et al., 2004; for a review, see Knutson and Cooper, 2005). These structures constitute the so-called reward system. Outcome-related activations have, however, also been found in other structures such as the insula (Clark et al., 2009; Delgado et al., 2000) and the hippocampus (Dickerson et al., 2011). Furthermore, a study on patients with cerebellar lesions suggests that the

* Corresponding author. Tel.: +49 234 32 23119; fax: +49 234 32 14622.
E-mail addresses: stefan.kobza@ruhr-uni-bochum.de (S. Kobza), christian.bellebaum@hhu.de (C. Bellebaum).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.10.036
0028-3932/© 2014 Elsevier Ltd. All rights reserved.

cerebellum is also involved in reward-based reversal learning (Thoma et al., 2008), which is in line with anatomical connections between the cerebellar dentate nucleus and the striatum (Hoshi et al., 2005). Consequently, the cerebellum may also play an important role in non-motor (stimulus–outcome) learning.

Importantly, reinforcement-learning is not necessarily restricted to processing of action–outcome related PEs: an action may result in different outcomes depending on the context or, in experimental terms, a preceding informative 'cue'. Notably, one can thus differentiate between two outcome PEs. An action-independent PE reflects the difference between the received outcome and the subjective value (SV) of the cue, which, just as the action value (AV, i.e. subjective value of the chosen action), changes based on outcome history. An action-dependent PE, on the other hand, reflects the difference between the received outcome and the AV. Both PEs appear to be differentially processed in the brain. O'Doherty et al. (2004) showed that the ventral striatum is involved both when outcomes did and did not depend on a preceding action, whereas the dorsal striatum codes action-dependent PEs. It is thus conceivable that especially the dorsal striatum facilitates learning of associations between (own) actions and their consequences. In line with this assumption, Bellebaum et al. (2008) reported disrupted feedback-based reversal learning in BG patients especially when the dorsal striatum was affected.

Stimulus–action–outcome associations can also be learned via observation of another person's actions and the feedback he or she receives. On the one hand, observational learning is characterised by an additional PE which relates to observed actions and which is coded in the dorsolateral PFC (Burke et al., 2010). Furthermore, Monfardini et al. (2013) found activations for observed but not own incorrect outcomes in the posterior medial frontal cortex, the anterior insula, and the posterior superior temporal sulcus. On the other hand, many brain regions are involved in processing of outcome PEs for both active and observational learning, with decreased activity of parts of the 'classical' reward system in observational as compared to active learning (Bellebaum et al., 2010, 2012; Yu and Zhou, 2006). In an fMRI study by Bellebaum et al. (2012), PE-dependent activations in the right putamen were found in both types of learning, with stronger activations in the right anterior caudate nucleus for active learners, suggesting that the processing of PEs is generally reduced in observational learning from feedback. On the other hand, these studies showed that action–outcome associations are learned similarly well in active and observational learning, thereby demonstrating that action–outcome associations can be acquired by observation. We hypothesised, however, that parts of the reward system are dedicated to integrating own (rather than observed) actions with outcomes, and we examined this by differentiating between PEs depending and not depending on the preceding (own or observed) action.

Based on our recent fMRI findings (Bellebaum et al., 2012) and evidence we obtained in PD patients (Kobza et al., 2012), we expected that the BG integrate own actions with outcome information during outcome evaluation in active learning, which would lead to differences between active and observational learning with respect to action-dependent PE processing. We further hypothesised that neural coding of outcome PEs in the reward system is not enhanced in active relative to observational learners if PEs are independent from actions.

For both SVs and AVs, activations have been found in parts of the reward system, such as the orbitofrontal cortex (FitzGerald et al., 2009), the dorsal ACC (Camille et al., 2011), the PFC (Glascher et al., 2009), the supplementary motor cortex (Wunderlich et al., 2009), and the putamen during active learning (FitzGerald et al., 2012). Activations reflecting reward expectation have so far not been investigated in observational learning. Furthermore, activity related to uncertainty has been reported for the amygdala in fMRI studies on aversive conditioning (Buchel et al., 1998; Labar et al., 1998) but also reward learning (Prevost et al., 2011). Uncertainty reflects the extent to which expectations of future reward vary over the course of the task (for the computational definition, see Section 2.4.4): Prior to learning, outcomes are completely unknown, so that uncertainty is at its maximum. Over the course of learning, outcome predictions become more accurate, so that uncertainty decreases. Consequently, uncertainty can be regarded as an inverse indicator of stimulus–outcome learning such as in classical conditioning, which has been shown to depend on the cerebellum (Daum et al., 1993). Thus, the present study also aimed to explore similarities and differences between active and observational learning with respect to the neural representation of SVs, AVs, and uncertainty signals preceding the outcome phase.

2. Material and methods

2.1. Subjects

33 healthy, right-handed adult volunteers participated in the study. Two participants were excluded due to data acquisition problems. Out of the remaining 31 subjects (age range 20–34 years), one group of 16 subjects (6 female; mean [M] age$=$25.1 years; standard deviation [SD]$=$3.7 years) engaged in an active feedback learning task, whereas the 15 subjects (9 female; $M=23.9$ years; $SD=4.5$ years) of a second group learned by observing the choices and following feedback in another person (see Section 2.2 for details of the learning tasks). The mean age did not differ between groups ($p=.45$). The current IQ as estimated via the Multiple Choice Vocabulary Test (Mehrfachwahl-Wortschatz-Test, MWT, version B; Lehrl, 2005) was also comparable ($p=.48$) between groups who learned actively ($M=116.3$; $SD=13.9$) or by observation ($M=119.8$; $SD=14.0$). All participants had normal or corrected-to-normal vision. Apart from standard exclusion criteria applied in fMRI studies – such as artificial cardiac pacemakers, metallic implants, diagnosed or reported claustrophobia – a history of neurological or psychiatric disease and regular medication affecting the central nervous system led to exclusion from the study. Prior to participation, subjects gave written informed consent. The study conforms to the Declaration of Helsinki and received ethical clearance by the Ethics Board of the Faculty of Psychology of the Ruhr University Bochum, Germany.

2.2. The learning tasks

In the present study, two feedback learning tasks were used: one in which subjects learned from own choices and the following outcomes, and one in which subjects learned from choices of another subject and the following outcomes. Both tasks are based on a probabilistic learning card-guessing paradigm introduced by Delgado et al. (2005). As in our previous studies on differences between active and observational learning (Bellebaum et al., 2010; Kobza et al., 2012), we applied a between-subjects design. This design rules out carry-over effects, which may occur in within-subject designs, in which learning from own choices in one phase may influence behaviour and/or neural coding of learning by observation in another phase of the experiment and vice versa.

Recording of participants' responses and timing of stimuli – presented via MRI video goggles (Resonance Technology, Inc.; http://www.mrivideo.com) – was controlled by Presentation Software (Neurobehavioral Systems, Inc.; http://www.neurobs.com).

2.2.1. Active feedback learning task

The subjects of the first group learned from their own choices. Each active learning trial started with the presentation of a fixation cross with a duration of 4000, 8000, and 12000 ms on 58.3%, 29.2%, and 12.5% of trials, respectively. Then one out of five different 'cards' (represented by frames including a circle, triangle, square, star, or hexagon) was presented as cue. Subjects were instructed that on the back of each card a number would be printed. After another fixation cross, subjects were asked to guess whether the number was lower or higher than the number 5, i.e. between 1 and 4 or between 6 and 9, by selecting a downward-directed arrow (presented on the left) or an upward-directed arrow (presented on the right) using the left or right button of a response box via index or middle finger of the right hand, respectively. Following the response, the chosen arrow was surrounded by a red circle. If subjects did not respond within 2000 ms, they were prompted to respond faster. Otherwise, after another presentation of a fixation cross, positive ('+50c' in green characters, indicating a monetary win of 50 cents) or negative ('−50c' in red characters, indicating a monetary loss of 50 cents) feedback was given for a correct or an incorrect guess, respectively (see Fig. 1A for the sequence of events in one learning trial and for the duration of stimulus presentation).

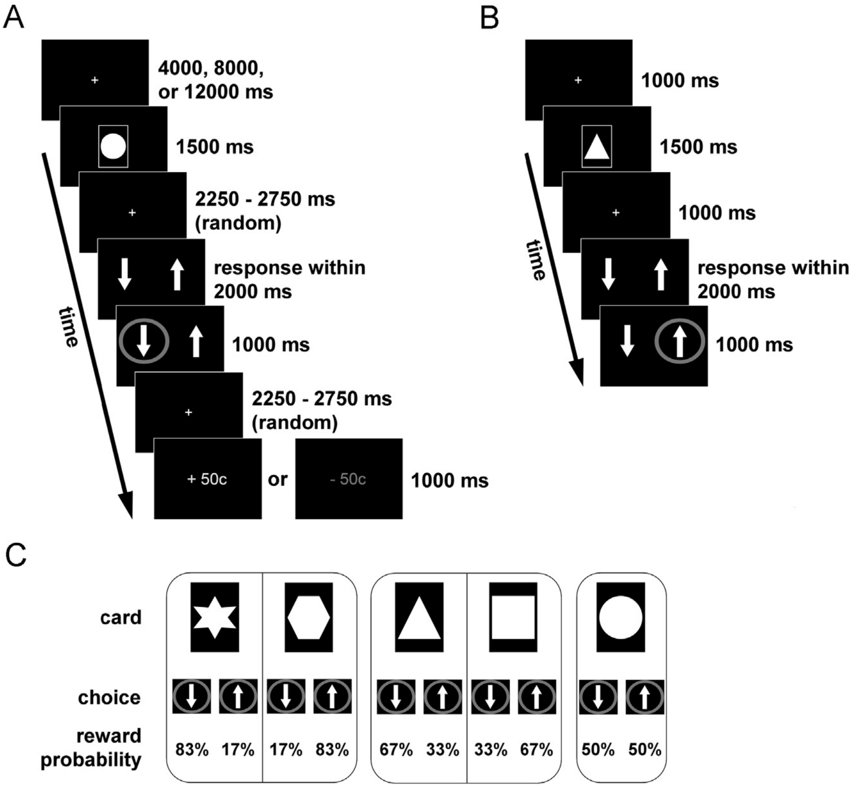

Fig. 1. The active and observational learning task. (A) Sequence of events in one (active or observational) learning trial and durations of stimuli. Subjects were presented one 'card' out of five and had to either make (active learning task) or observe (observational learning task) a guess on whether a digit on the back of the card was lower or higher than 5 as indicated by a red circle surrounding the left or right arrow, respectively. Shortly after, positive or negative monetary feedback was given. (B) Time course of events in one test trial. Test trials were identical for both groups, and no feedback was given. (C) 'Card' stimuli and associated reward ('+ 50c') probabilities following each choice. Note that both choices were equally correct or incorrect for the 'card' associated with 50% reward probability, because irrespective of choice, wins and losses occurred equally often for this 'card'.

Participants were told that they could improve their guesses and thus maximise their monetary gain over the course of the experiment by learning from previous choices and associated outcomes separately for each card. Unbeknownst to the subjects, the reward probabilities of the five cards were 83% (two cards: correct guess high or low), 67% (two cards: correct guess high or low) or 50% (one card: equal frequencies of positive and negative feedback irrespective from guesses). For incorrect responses, reward probabilities are reversed, e.g. for cues which lead to positive feedback in 83% following correct guesses, wrong guesses lead to positive feedback in the remaining 17%. Thus, on each trial, one guess led to positive feedback, while the other guess led to negative feedback (see Fig. 1C for choices and reward probabilities associated with each card).

The paradigm also involved test trials, which were identical with the learning trials except that no feedback was given following guesses, and the duration of fixation crosses before and after card presentation was shortened to 1000 ms each. Test trials served to provide a comparable assessment of learning in the active and observational (see Section 2.2.2) learning tasks. Subjects were asked to use the knowledge previously gained in learning trials to make optimal guesses on test trials. €20 was guaranteed as compensation for volunteering, and subjects were informed that they could receive a monetary bonus for outstanding performance on both learning and test trials (see Fig. 1B for the sequence of events in one test trial and for the duration of stimulus presentation).

Each of two blocks of 60 learning trials was followed by a block of 60 test trials, resulting in four blocks in alternating order. In every block, each card was presented twelve times.

2.2.2. *Observational feedback learning task*

In the observational learning task, each subject had to learn by observing the guesses of and outcomes for another subject who previously performed in the active learning task. In order to create a realistic observation scenario, each observer drew an ID from a box to determine the active subject to be observed. It was emphasised that observed guesses and outcomes were not simulated or generated by a computer but recorded from the corresponding subject.

In blocks of learning trials, each subject observed all stimuli exactly as presented to the observed subject (see Fig. 1A), leading to matched experimental variables in pairs of subjects who learned actively or by observation. To match the motor requirements of the active learning task, after each presentation of arrows, observers had to press a button in order to see the observed subject's choice, as subsequently indicated by a red circle surrounding the chosen arrow, and the outcome the observed subject received for his or her choice. To exclude mismatches between the sides of the pressed button and the observed choice, observers had to press the same button on each learning trial, which was different from the two buttons used by the active learners. If observers did not press the button within 2000 ms, they were prompted to respond faster.

As in the active learning task, each of two blocks of 60 learning trials was followed by a block of 60 test trials (see Fig. 1B). The blocks of test trials were identical for both learning tasks, that is, observers also had to make active choices in test trials to allow between-group comparison of learning. To exclude learning from outcomes of own choices, no feedback was given in test trials.

Importantly, to ensure that the significance of outcomes was comparable between groups, observers were told that the amount of money paid out at the end of the experiment depended both on observed performance in learning trials and own performance in test trials. €20 was guaranteed as compensation for volunteering.

2.3. *Procedure*

Subjects were informed that the study purpose was the investigation of brain mechanisms of active and observational learning from feedback. After exclusion criteria had been ruled out and written consent had been given, subjects were placed in the MRI scanner, and either the active or the observational learning task was started. Following the learning task, compensation for volunteering was paid out.

2.4. *Data analysis*

2.4.1. *Behavioural data analysis*

Choice accuracy was analysed by means of repeated-measures ANOVAs. Choices were classified as correct if the associated reward probability was higher as compared to the alternative choices. For the cue which yielded equal frequencies of wins and losses of 50 cent independent from guesses (reward probability of 50%), we deliberately defined guesses of numbers higher than 5 as correct responses.

In active learners, an ANOVA involving the within-subjects factors BLOCK (1 vs. 2) and PROBABILITY (83% vs. 67% vs. 50%) was performed on accuracy during learning trials. Between-group comparisons were based on choices in test trials.

Performance was analysed by means of an ANOVA with BLOCK and PROBABILITY as within-subjects factors and GROUP (active vs. observation) as between-subjects factor.

As all subjects were instructed to learn from feedback, observers were expected not to imitate active learners' choices but to learn from feedback to observed choices. For learning from feedback, observers' test performance was expected to show only weak correlations with active learners' performance observed in the preceding learning block: a high-performing observer would learn from feedback to both correct and incorrect choices of a low-performing active learner, leading to higher performance of the observer in test blocks as compared to the observed active learner's performance in learning blocks. In turn, a low-performing observer would show poor learning from feedback to both correct and incorrect choices of a high-performing active learner, leading to lower performance of the observer in test blocks as compared to the observed active learner's performance in learning blocks. In contrast to observers, active learners were expected to show high correlations between learning and test block performance, because their performance was expected to be high or low throughout the task for both learning and test blocks, that is, active learners would carry over high and low performance in active learning blocks to high and low performance in subsequent test blocks, respectively.

Taken together, correlations between mean (own or observed) performance on learning blocks and performance on test blocks were calculated, separately for each group and for the first and second half of the task, to (1) verify that performance in test blocks was a valid measure of learning, as would be indicated by significant correlations for the group of active learners, and (2) rule out that observers merely imitated the choice behaviour of the observed subjects, which would lead to significant correlations for the group of observational learners (note, however, that significant correlations would not necessarily indicate imitation learning, e.g. if performance in learning from feedback is comparable in pairs of active and observational learners). Consequently, we expected significant correlations between performance in learning and test blocks for active learners only, and significantly lower correlations in observers. As learning was only possible for the 67% and the 83% conditions, data from the 50% condition did not enter correlational analyses on performance.

The level of significance was set to $p < .05$ (two-tailed) for all behavioural data analyses; in correlation analyses, Bonferroni corrections were applied. When the sphericity assumption was violated, the Greenhouse–Geisser correction to adjust the degrees of freedom was applied. To resolve interactions whenever necessary, post-hoc t tests (two-tailed) were performed. PASW Statistics 18 (SPSS Inc., Chicago, IL, USA) was used for all behavioural analyses.

2.4.2. fMRI data acquisition

MR Images were acquired on a 3 T MRI scanner (Achieva, Philips Healthcare, Einthoven, The Netherlands) with a 32-channel SENSE head coil. First, a high resolution 3D T1-weighted structural scan was acquired for each subject (echo time (TE) 3.8 ms, flip angle 8°, FOV 240 mm × 240 mm, 220 slices, 1 mm × 1 mm × 1 mm voxel size). Separately for each of the two learning blocks, a sequence of echo-planar images (EPI) with a TE of 35 ms and a repetition time (TR) of 2400 ms was acquired, with 34 slices (no gap, slice thickness = 4 mm, in-plane resolution 2 mm × 2 mm, for reconstructed voxels: 1.65 mm × 1.65mm) per whole brain volume and 400 whole brain volumes per block, yielding a total of 800 volumes per subject. Five additional dummy volumes at the beginning of each sequence were discarded to allow for BOLD signal stabilisation.

In order to gain whole brain volumes including the cerebellum completely, no additional tilt was applied for EPI scans.

2.4.3. fMRI data analysis

SPM8 (Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, London, UK) was used for pre-processing and statistical analysis of fMRI data. EPI were slice-time corrected and realigned relative to the first acquisition sequence. Following realignment, EPI were unwarped (Andersson et al., 2001), co-registered with the same subject's structural images, spatially normalised to the Montreal Neurological Institute (MNI) standard space, resampled to 2 mm × 2 mm × 2 mm voxel size, and smoothed by a 6 mm full-width at half-maximum (FWHM) Gaussian kernel. Then, General Linear Model statistical analysis was used (Friston et al., 2002). A two-stage random-effects approach was adopted to ensure generalisability of the results at the population level (Penny and Holmes, 2003). The time series of each participant were high-pass filtered at 128 s. No global normalisation was performed.

The analysis aimed at identifying those brain regions which were significantly modulated by (1) the SV of a cue (card), (2) the uncertainty associated with a cue, the AV following (3) cue presentation and (4) response, the outcome-related PE depending on (5) the SV of the cue and (6) the response-related AV. For this purpose, for all (learning) trials across the two sequences of fMRI data acquisition, fMRI data of each subject were modelled with three cue regressors (comprising the 5 card stimuli), where the size of the SV, the uncertainty, and the difference between the two AVs ($AV_{right} - AV_{left}$), as suggested by FitzGerald et al. (2012), in each individual trial were used as parametric modulators. Furthermore, one response regressor corresponded to the guesses (number lower or higher than 5), with the AV difference being used as parametric modulator. Additionally, two regressors relating to the outcome (win or loss of 50 cent) were introduced, one of which was parametrically modulated by the PE depending on SV of the cue, and the other depending on the AV of the performed or observed response. Finally, regressors of no interest, referring to the onsets of fixation crosses and arrows, were added to the model.

Linear contrasts were brought to the group level for all parametric modulators separately by means of one-sample t tests. For each parametric modulator, a factorial design including the between-subject factor GROUP (active vs. observational learners) was specified on the group level based on the first level t contrasts. This analysis served to explore activations related to the parametric modulators both separately for each group and in between-group difference analyses.

All reported statistics refer to whole brain analyses. Statistical maps were thresholded at $p = .01$, corrected for multiple comparisons using cluster-size thresholding based on Monte Carlo simulation (10 000 runs; Slotnick et al., 2003), with a threshold of $p = .001$ for single voxels, yielding an extent threshold of 25 contiguous voxels. This approach was chosen to reduce the type II error, i.e. missing true effects, while keeping the type I error, i.e. false alarms, at a reasonably low level, as was suggested by Lieberman and Cunningham (2009). Activation foci coordinates are reported in MNI space and were transformed to Talairach space (see http://imaging.mrc-cbu.cam.ac.uk/downloads/MNI2tal/mni2tal.m) via non-linear transformation for anatomical labelling using the stereotaxic atlas of Talairach and Tournoux (1988). We were particularly interested in modulations of activity in regions previously reported to be involved in outcome or error processing (see Section 1). These regions of interest (ROIs) consisted bilaterally of the putamen, the caudate nucleus, the globus pallidus, the ACC, the hippocampus including parahippocampal gyrus, the cerebellum, and the PFC.

2.4.4. Calculation of parametric modulators

In accordance with a model introduced by Prevost et al. (2011), the SV of the cue V was updated on each trial t, separately for each subject and cue s (cards 1–5), as expressed by the following equation:

$$V_s(t+1) = V_s(t) + \delta \times \sqrt{U(t)/U(1)} \times (R(t) - V_s(t)) \qquad (1)$$

$R(t) - V_s(t)$ represents a PE, i.e. the deviation between SV V_s and outcome R, with $R = 1$ for win and $R = -1$ for loss of 50 cent, on trial t. The update on trial $t+1$ is performed by weighting the PE with the learning rate δ. The trial-by-trial uncertainty $U(t)$ (see Eq. (2)), representing the variation of reward expectation over time, was used to adjust the learning rate, with $\delta = 0.2$, in an approximately optimal manner (Prevost et al., 2011). Apart from the adjustment of learning rate for SVs, uncertainty $U(t)$ also served as the second cue-related parametric modulator. In accordance with Prevost et al. (2011), a reinforcement learning-based approximation of a Kalman filter was adopted to compute $U(t)$, which served as a model-based prediction of uncertainty and decreased over the course of learning as variance of outcome predictions decreased. $U(t)$ was updated according to

$$U(t+1) = U(t) + \delta \times ((V_s(t+1) - (V_s(t))^2 - U(t)) \qquad (2)$$

As was pointed out by Prevost et al. (2011), the trial-by-trial uncertainty $U(t)$ is updated by a weighted difference between the squared deviation between the previous and the current SV and the previous uncertainty value.

To estimate the learning rates for AVs and outcome PEs (see below), we first assumed a subjective value Q of 0 cent for each cue at the beginning of the experiment, i.e. Q at trial $1 = Q(1) = 0$ cent. Next, we calculated Q separately for each subject, cue, and end of learning block by means of analysing the choices in the subsequent test block both for active and observational learners to provide comparable AV and PE calculations in the two groups. For each cue, Q was calculated as the sum of the proportion of correct responses times $+50$ cent and proportion of incorrect responses times -50 cent, e.g. after the first learning block, a rate of 100% correct or incorrect responses in the following test block indicated Q at trial $60 = Q(60) = +50$ cent (100% $+50$ cent outcomes plus 0% -50 cent outcomes) or $Q(60) = -50$ cent (0% $+50$ cent outcomes plus 100% -50 cent outcomes), respectively, with a linear relation between test block accuracy and SV.

As was done by Bellebaum et al. (2012), learning rates were determined for each subject, block, and cue separately, so that

$$Q_s(t+1) = Q_s(t) + \alpha_s \times (R(t) - Q_s(t)), \qquad (3)$$

a simple delta rule learning model (Gluck and Bower, 1988), fit the known subjective values at the beginning and end ($Q(1)$, $Q(60)$, and $Q(120)$) of each learning block in an optimal manner. If, for example, $Q(60) = 15$ cent increased to $Q(120) = 35$ cent, then the learning rate was chosen such that the sequence of positive and negative feedback following that cue resulted in this increase by 20 cent.

By means of α, we calculated the AVs according to a fictive reinforcement learning model (Fudenberg and Levine, 1998; Myung and Busemeyer, 1992; note that reinforcement learning models based on feedback to own choices and, thus, comparisons regarding model fits could not be considered in the present study, as only active learners received feedback to own choices). In contrast to simple and decay reinforcement learning models, the fictive model accounts for the reversal structure of outcomes available on each trial of the present learning tasks: every monetary win or loss following a performed or observed action implies that the inverse outcome, i.e. a monetary loss or win, respectively, would have followed the alternative action. Consequently, the AVs for both selected and unselected actions were updated on each learning trial t according to the following equation, in which Q represents the value of action j, separately for each block, subject, and cue s:

$$Q_{s,j}(t+1) = Q_{s,j}(t) + \alpha_s \times (R(t) - Q_{s,j}(t)) \tag{4}$$

If j is the selected (performed or observed) action, $R(t)$ is the outcome (50 for win and -50 for loss of 50 cent) received on the current trial. If j is the unselected action, $R(t)$ is the inverse of the received outcome, with -50 for win and 50 for loss of 50 cent. The AV of the selected or unselected action at trial $t+1$ is the sum of the AV and, weighted by the learning rate α_s, the PE between received or inverse outcome R and AV at the previous trial of the same cue.

Notably, different outcome-related PEs were modelled in the last two equations. $R(t)-Q_s(t)$ represents the deviation between the subjective value Q of a cue and the outcome R obtained on a particular trial t, independent from an action. $R(t)-Q_{s,j}(t)$, however, is an action-dependent PE, referring to the deviation between the value Q of action j and outcome R. Both of the PEs served as separate outcome-related parametric modulators, with the action-dependent PE only referring to selected actions.

3. Results

3.1. Behavioural data

Active learners' mean accuracies for each probability and learning block as well as learning curves over the course of trials are depicted in Fig. 2A. ANOVA on accuracy in learning trials yielded a significant main effect of PROBABILITY ($F_{[2,\ 30]}=3.682$; $p=.037$), reflecting higher accuracy for cues with a reward probability of 83% as compared to 67% and 50%, and a near-significant main effect of BLOCK ($F_{[1,\ 15]}=4.241$; $p=.057$), with higher accuracy in the second as compared to the first block. The interaction between PROBABILITY and BLOCK did not reach significance ($p=.252$). Active learners' performance significantly differed from chance level, that is, 50% correct choices, for cues with a reward probability of 83% in the first ($t(15)=5.146$; $p<.001$) and second ($t(15)=5.724$; $p<.001$) learning block. For cues with a reward probability of 67%, a trend in the first learning block ($t(15)=1.804$; $p=.091$) was followed by a significant difference from chance level in the second learning block ($t(15)=2.245$; $p=.040$). For cues with a reward probability of 50%, active learners made significantly more right-choices than expected from chance level, that is, 50% left- and 50% right-choices, both in the first ($t(15)=2.158$; $p=.048$) and second ($t(15)=4.388$; $p=.001$) learning block. For the 67% and the 83% conditions, over the course of learning and test blocks, active learners' performance was above chance level both for cards associated with correct left ($t(15)=5.307$; $p<.001$) and correct right ($t(15)=3.831$; $p=.002$) choices, both of which did not differ in frequency ($p=.333$).

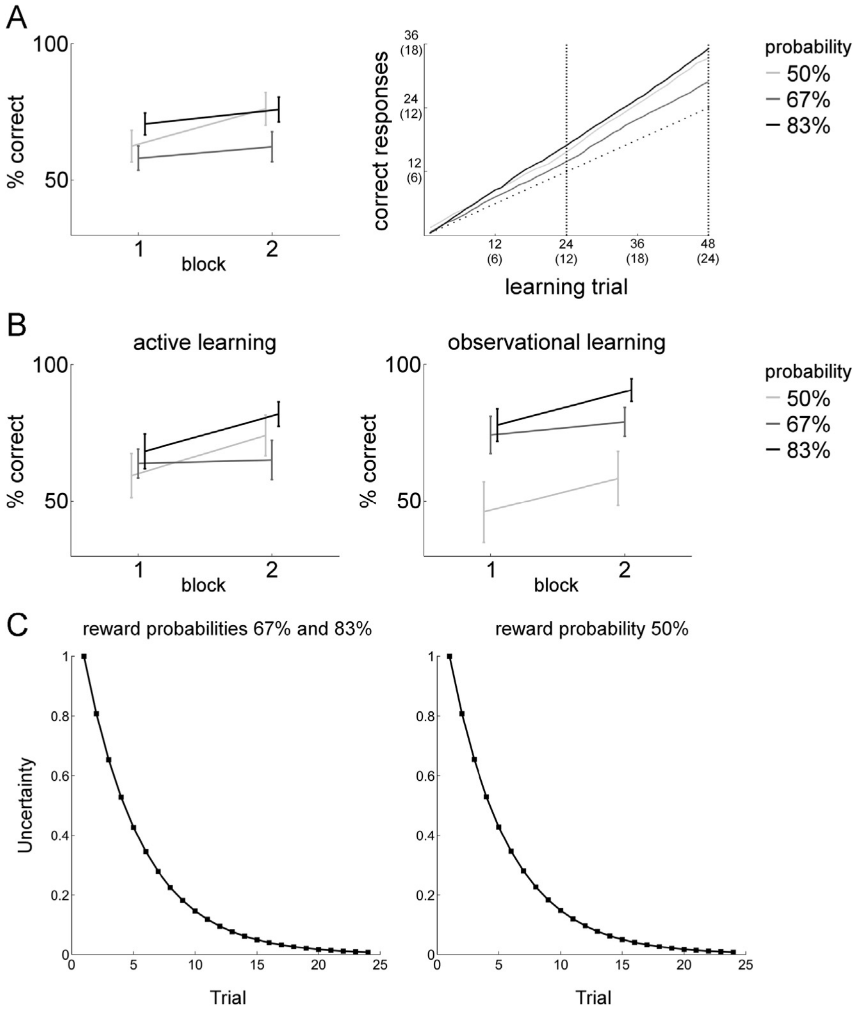

Fig. 2. Learning performance for each probability. (A) Active learners' performance in learning trials including feedback. Left: mean performance in blocks 1 and 2. Right: mean accumulated number of correct responses over the course of learning trials. Numbers in brackets refer to 50% probability condition, for which each block included 12 instead of 24 trials. The end of the first and second learning block is indicated by vertical dotted lines at trial number 24 (12) and 48 (24), respectively. The dotted curve indicates chance performance. (B) Mean performance of active (left) and observational (right) learners in blocks of test trials without feedback. (C) Estimation uncertainty averaged across all subjects and learning blocks. Left: uncertainties averaged across cues with reward probabilities of 67% and 83% (non-chance; uncertainty curves did not visually differ between separate cues). Right: uncertainties separately for the cue with reward probability of 50% (chance). Note that curves show an approximately exponential decay over non-contiguous trials, because trial types were interleaved. Error bars indicate standard errors of the mean.

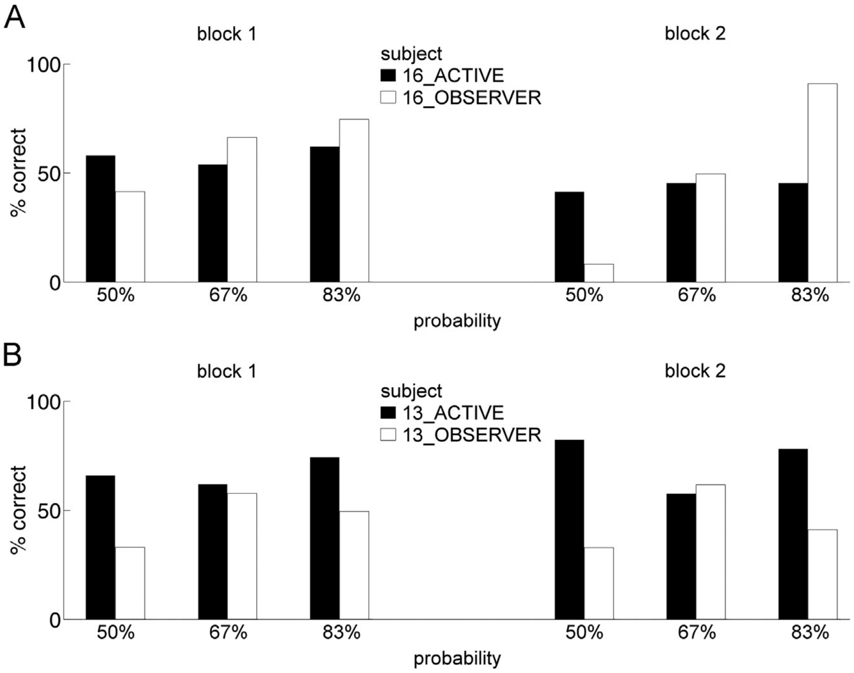

Fig. 3. Sample learning trial performances of active learners and test trial performances of observers. (A) Subject 16_OBSERVER shows higher performance in test trials than 16_ACTIVE in learning trials (67% and 83% probability). (B) Subject 13_OBSERVER shows lower performance in test trials than 13_ACTIVE in learning trials (except for 67% probability in block 2).

In the ANOVA on accuracy in test trials, a significant main effect of BLOCK emerged ($F_{[1,\ 29]}=6.452$; $p=.017$), resembling the pattern in active learning trials (see above), that is, higher accuracy in the second as compared to the first block. Furthermore, a significant main effect of PROBABILITY was found ($F_{[2,\ 58]}=8.358$; $p=.002$): Accuracy was higher for cues with a reward probability of 83% as compared to 67% ($t(30)=2.629$; $p=.013$) and 50% ($t(30)=3.723$; $p=.001$). A significant interaction between PROBABILITY and GROUP ($F_{[2,\ 58]}=4.346$; $p=.028$) was driven by more right- than left-choices for 50% reward probability in active as compared to observational learners, whereas the opposite pattern was found for 67% and 83%. The corresponding *t* tests, however, did not yield significant differences (all $p>.104$). In the ANOVA, no further main effects or interactions emerged (all $p>.237$). For cues with a reward probability of 83%, performance differed significantly from chance level for active learners in the first ($t(15)=2.885$; $p=.011$) and second ($t(15)=7.166$; $p<.001$) test block and also for observational learners in the first ($t(14)=4.684$; $p<.001$) and second ($t(14)=9.853$; $p<.001$) test block. Along similar lines, performance was higher than chance level for cues with a reward probability of 67% in active learners' first ($t(15)=2.641$; $p=.019$) and second ($t(15)=2.132$; $p=.050$) test block and in observational learners' first ($t(14)=3.557$; $p=.003$) and second ($t(14)=5.506$; $p<.001$) test block. For cues with a reward probability of 50%, active learners did not make significantly more right-choices than expected from chance level in the first ($p=.260$) but in the second ($t(15)=3.233$; $p=.006$) test block, whereas in observational learners, comparisons against chance level yielded no significant differences in either of the two test blocks for this cue (both $p>.41$). Fig. 2B illustrates mean accuracies on test blocks separately for each group, probability, and block. Uncertainties over the course of learning trials are depicted in Fig. 2C.

In active learners, performance correlated significantly between the first learning and test block, and between the second learning and test block ($r=.721$; $p=.002$, and $r=.783$; $p<.001$, respectively). In observational learners, however, both the correlation between observed performance in the first and second learning block, and active performance in the first and second test block, respectively, were not significant (both $p>.615$). Following a Fisher *z* transformation, between-group comparisons revealed significant differences in the strength of the correlations: the correlation between (active or observed) performance in learning and active test trials was higher for active learners as compared to observers both in the first ($z=1.99$; $p=.023$) and the second ($z=2.28$; $p=.011$) half of the task. Sample performances of high- and low-performing observers and low- and high-performing observed active learners, respectively, are shown in Fig. 3.

In order to address potential multicollinearity between the terms derived from the learning model, we also calculated variance inflation factors (VIF) separately for each parametric modulator. The highest VIF emerged for the action-independent PE ($VIF=3.024$; all remaining $VIF<1.02$), which is acceptably low (Myers, 1990; Menard, 1995).

3.2. fMRI data

3.2.1. Parametric modulation of activity by SV of the cue

For active learners, a significant modulation was found in the right medial frontal gyrus. Further analyses separately in observational learners and between-group comparisons (active > observational or observational > active) yielded activations only in clusters out of ROIs (Supplementary Table 1 lists all activations related to the SV of the cue).

3.2.2. Parametric modulation of activity by uncertainty associated with the cue

With respect to cue-related uncertainty, large clusters emerged bilaterally in the cerebellum (declive) for observational learners only. Additional modulations were found in the right cerebellar culmen, and in the left and right middle frontal gyrus, the latter of which showed stronger modulation in observational as compared to active learners. In the opposite contrast, i.e. for active as compared to observational learners, larger modulations were

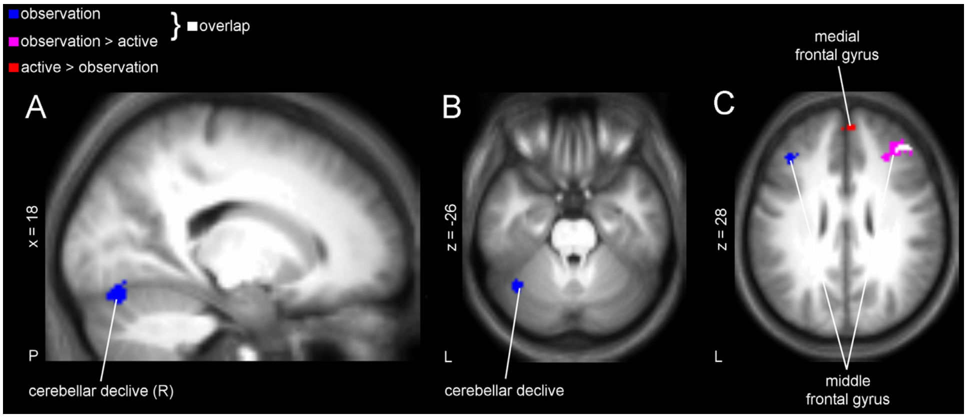

Fig. 4. Uncertainty-related activations in active and observational learners. For observational learners (*blue*), activations were found in the right (A) and left (B) cerebellum (see Supplementary Table 2 for coordinates of local activation peaks). (C) Activity in the middle frontal gyrus was significantly modulated in observational learners (peak activations left hemisphere: $x=-36$, $y=30$, $z=30$; $Z=4.16$; right hemisphere: $x=42$, $y=36$, $z=26$; $Z=3.96$) and stronger as compared to active learners (*white*) in the right middle frontal gyrus ($x=36$, $y=36$, $z=26$; $Z=4.65$), exceeding (*pink*) the area activated in observational learners only. The opposite pattern, that is, stronger activity in active than observational learners (*red*), emerged for the right medial frontal gyrus ($x=4$; $y=50$; $z=28$; $Z=3.80$; see Supplementary Table 2 for activations out of the ROIs).

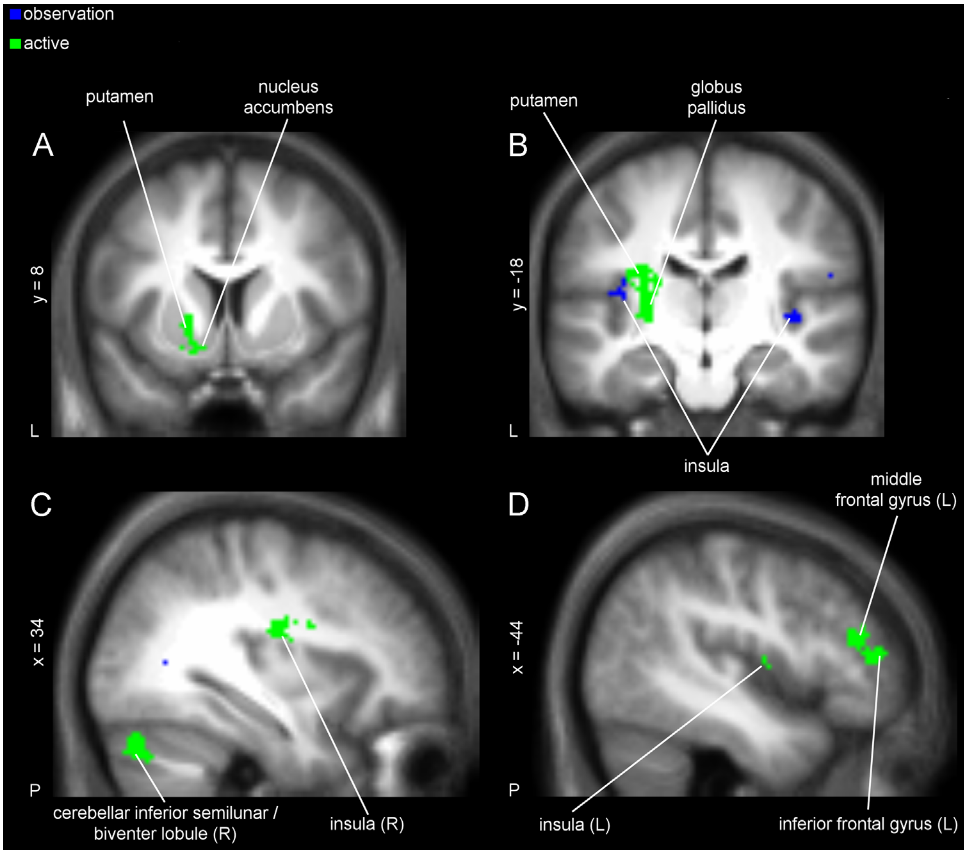

Fig. 5. Activations related to action-independent outcome PEs in active and observational learners. Separately for active learners (*green*), activity emerged for (A) the left putamen (peak activity at $x=-18$, $y=8$, $z=0$; $Z=4.58$), extending to the nucleus accumbens, (B) the left globus pallidus ($x=-26$, $y=-16$, $z=4$; $Z=4.51$), (C) the right cerebellum*, and (D) the left middle* and inferior ($x=-44$, $y=42$, $z=12$; $Z=4.06$) frontal gyrus. For both observational (*blue*); (B) and active (C and D) learners, activations emerged bilaterally in the insula* (see Supplementary Table 5 for activations out of the ROIs). Note that slices in (B) and (C) were selected to simultaneously visualise multiple clusters of activations, thereby showing globus pallidus and insula activations, respectively, extending to the white matter. Importantly, these activations peak in the grey matter. (* See Supplementary Table 5 for coordinates of local activation peaks).

observed in the right medial frontal gyrus (see Fig. 4 and Supplementary Table 2). No modulations were found in ROIs when active learners were considered separately. Thus, whereas the cerebellar culmen and declive, and the middle frontal gyri play a role in observational learners only, the right medial frontal gyrus was more strongly involved in active learners.

3.2.3. *Parametric modulation of activity by AV following cue presentation*

For activity related to the AV following cue presentation, between-group comparisons showed no clusters of significant modulation. Along similar lines, the separate analysis for observational learners yielded no suprathreshold activation clusters. For active learners,

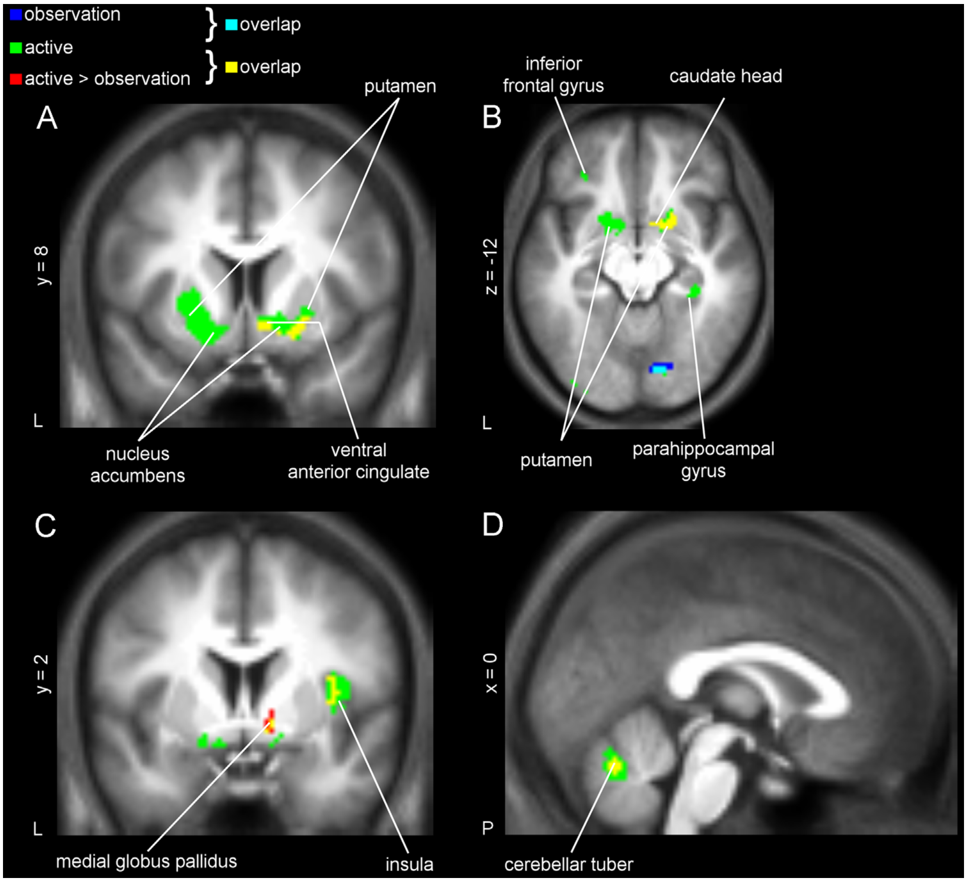

Fig. 6. Activations related to action-dependent outcome PEs in active and observational learners. For observational learners (*blue*), no activations in ROIs were found. For active learners (*green*), however, activations emerged (A) bilaterally for the putamen*, extending to the nucleus accumbens and the ventral anterior cingulate, (B) the left inferior frontal gyrus (peak activity at $x = -30$, $y = 34$, $z = -10$; $Z = 3.72$), the right parahippocampal gyrus*, (C) the right insula ($x = 38$, $y = 2$, $z = 12$; $Z = 4.89$), and (D) bilaterally for the cerebellum*. Furthermore, activity in (A) the right ventral anterior cingulate ($x = 6$, $y = 8$, $z = -10$; $Z = 4.15$), (B) the right putamen ($x = 16$, $y = 6$, $z = -12$; $Z = 3.95$), extending to the caudate head, (C) the right medial globus pallidus ($x = 12$, $y = 2$, $z = -4$; $Z = 4.40$), the right insula*, and (D) bilaterally in the cerebellum (left hemisphere: $x = -6$, $y = -60$, $z = -28$; $Z = 3.60$; right hemisphere: $x = 0$, $y = -66$, $z = -24$; $Z = 3.90$) show (also; *yellow*) stronger modulations (*red*) as compared to observational learners (see Supplementary Table 6 for activations out of the ROIs). Note that the slice in (B) was selected to simultaneously visualise multiple clusters of activations, thereby showing globus pallidus/putamen activations extending to the white matter. Importantly, these activations peak in the grey matter. (* See Supplementary Table 6 for coordinates of local activation peaks).

however, significant modulations emerged in the left middle frontal gyrus and insula (see Supplementary Table 3).

3.2.4. *Parametric modulation of activity by AV following response*

In contrast to AV-related modulations following cue presentation, significant modulations following responses were found in the left putamen for active learners (see Supplementary Table 4). Whereas the absence of this modulation in observers suggests that the left putamen is more strongly involved in active learners, between-group comparisons yielded no suprathreshold activation clusters.

3.2.5. *Parametric modulation of activity by outcome-related PE depending on SV of the cue*

In active learners, significant modulations emerged for the left BG (putamen and lateral globus pallidus), middle and inferior frontal gyrus, bilaterally for the insula, and for the right cerebellum (pyramis, declive, and uvula). In observational learners, modulations were found bilaterally in the insula (see Fig. 5 and Supplementary Table 5). Between-group comparisons yielded no suprathreshold activation clusters for either direction.

3.2.6. *Parametric modulation of activity by outcome-related PE depending on response-related AV*

Separate analyses revealed that, whereas no modulations in ROIs were found for observational learners, significant modulations emerged for the right parahippocampal gyrus, the left middle, inferior and medial frontal gyri, and, in large clusters, bilaterally for the cerebellum (bilaterally declive, left culmen and anterior lobe), the putamen, the right caudate head and insula in active learners. The between-group comparisons confirmed that the right BG (medial globus pallidus and putamen), anterior cingulate, insula, and bilaterally the cerebellum (right declive and left fastigium) show significantly stronger modulations in active as compared to observational learners (see Fig. 6 and Supplementary Table 6), whereas no stronger modulations emerged for observational as compared to active learners.

4. Discussion

The present study aimed to explore the neural correlates of active as compared to observational learning from feedback, focusing on stimulus–outcome and stimulus–action–outcome learning. To this end, we applied a task that involved learning of stimulus–response–outcome associations in a between-subjects design: for each cue stimulus, subjects of both groups could learn which of two actions was more likely to be followed by a monetary gain. Whereas one group learned from feedback to own actions, the second group learned from feedback to observed actions. Importantly, the separation of the cue and the action allowed us to analyse cue- and action-related activations independently, and, consequently, activations related to a cue-dependent (action-independent) and an action-dependent outcome PE.

4.1. The reward system and processing of action-dependent and action-independent PEs

In contrast to previous studies, the main focus of the present study was on processing of two different outcome PEs, an action-dependent PE referring to the difference between AVs and outcomes, driving (stimulus–)action–outcome learning as in instrumental conditioning, and an action-independent PE referring to the difference between SVs and outcomes, driving stimulus–outcome learning as in classical conditioning. For both PE types, active learners showed modulations peaking in the dorsal but also extending to the ventral striatum, in which PE activity is most consistently found (e.g. O'Doherty et al., 2004; Niv et al., 2007). Only for action-dependent PEs, however, striatal activation was stronger in active as compared to observational learners. In line with these results, Bellebaum et al. (2012) demonstrated in a recent study that the dorsal striatum is more involved in the coding of PEs in active as compared to observational learners. This finding was linked to a stronger involvement of the dorsal striatum in instrumental conditioning, in which actions are required to obtain an outcome, in contrast to classical conditioning (O'Doherty et al., 2004). The present study goes beyond this finding by demonstrating that differences in striatal activation are not caused by the learning type (active vs. observation) per se but rather by the strength of the link between outcome and (own) action: whereas activations in the dorsal striatum, particularly the right putamen and globus pallidus, were stronger for action-dependent PEs in active learners, no differences between groups were found for action-independent PEs. At first sight, these findings contradict a report on observers' goal-directed (response–outcome) learning depending on the dorsal anterior caudate more than the posterior caudate, which was more active in habitual (cue-response) learning (Liljeholm et al.; 2012). Note, however, that decreased striatal activity for action-dependent PEs in the present study refers to a comparison with active learners rather than another type of learning. Taken together, the functional dissociation between ventral and dorsal striatum seems comparable in active and observational learning, but striatal involvement is still reduced for action-dependent PEs in observational relative to active learning. The present study also adds to a previous study on coding of action and outcome PEs in observers' dorsolateral and ventromedial PFC, respectively (Burke et al., 2010), by demonstrating that the coding of outcome PEs is further influenced by the degree to which the outcome depended on an (own) action.

Action-dependent and action-independent PEs seem related to ‘actor’ and ‘critic’, respectively, in actor–critic models (e.g. Houk et al., 1995). Such models propose that a ‘critic’ updates predictions of future reward based on temporal difference PEs for stimulus-reward learning such as in classical conditioning, whereas the ‘actor’ uses a similar signal for stimulus–response(–reward) learning such as in instrumental conditioning (for a comparison on actor–critic models and a discussion on their plausibility in basal ganglia integration, see Joel et al., 2002). At first sight, stronger activations in active learners suggest that action-dependent PEs are an entity of the ‘actor’, because active learners had to make choices during learning trials, whereas observers merely saw active learners' choices and outcomes. Importantly, however, both groups had to learn associations between stimuli, (active or observed) responses and outcomes during learning trials. Therefore, between-group differences in activations related to action-dependent PEs do not necessarily indicate that they are an entity of the ‘actor’. Along similar lines, at first sight, comparable activations among groups suggest that action-independent PEs are an entity of the ‘critic’, because the ‘critic’ is independent from actions. Note, however, that each learning trial included (active or observed) choices, so that action-independent PEs in the present study are not necessarily an entity of the ‘critic’.

Note that low and high performers of both groups could not account for differences in the fMRI results for two reasons. First, the paradigm of our study allowed us to calculate learning rates based on test trial performance separately for each subject, block, and cue. This way, individual low and high learning rates entered the learning model for low- and high-performers, respectively, to yield an optimal calculation of PEs separately for each subject, block, and cue. Second, between-group differences in PE-related activity cannot be explained by between-group differences in performance per se, as subjects of both groups learned associations between cues, actions, and outcomes equally well, as was indicated by performance on test trials without feedback. The fact that the dorsal striatum showed stronger modulation by action-dependent PEs in active learning can thus not be related to between-group differences in prediction accuracy. In line with this finding, recent studies suggest that activity in the dorsal striatum relates to action execution rather than learning of action–outcome associations (Guitart-Masip et al., 2012; Shiner et al., 2012). The presence or absence of actions per se, however, cannot account for the differences between groups in the present study: motor requirements were identical for both groups, that is, all subjects were required to press a button in order to see the selected choice and following outcome. Therefore, the present findings rather suggest that the integration of action- and feedback-related information in the dorsal striatum requires the feedback to depend on *own* choices.

Recent studies suggest that activity of the dorsal striatum in feedback-based learning tasks is related to the execution rather than learning of goal-directed actions (Shiner et al., 2012; Smittenaar et al., 2012). In another recent fMRI study, Guitart-Masip et al. (2012) showed that activity in anticipation of action execution (go) as contrasted to action inhibition (no-go) differed in the SN, VTA, and dorsal striatum. Furthermore, for anticipation of reward, treatment with levodopa led to even increased activity in these structures for the go > no-go contrast, while reaction times decreased (Guitart-Masip et al., 2012). Taken together, stronger modulation of dorsal striatal activity for action-dependent PEs in active learners most likely refers to differences between groups with respect to action representations, probably modulated by the DA level. Consequently, a reduction in DA as in unmedicated PD patients affects active but not observational learning from feedback (Kobza et al., 2012), where the dorsal striatum plays a minor role (Bellebaum et al., 2012).

Learning from errors can be predicted from activity of the medial frontal cortex, as was shown both in an fMRI study (Hester et al., 2008) and as suggested by a study on the Feedback-related Negativity (van der Helden et al., 2010), which is an event-related potential component following negative feedback, and whose neural generator has been localised in the ACC (Gehring and Willoughby, 2002). With respect to the present study, higher anterior cingulate activity in active as compared to observational learners for action-dependent but not action-independent PEs may be related to previous findings of co-activation of the striatum and the ACC (Ridderinkhof et al., 2004; Rogers et al., 2004). Along similar lines, the ACC has not only been shown to play a role in error detection and correction (Casey et al., 1997; Garavan et al., 2002; Ullsperger and von Cramon, 2003) but also in action-related learning, which is impaired in patients with ACC lesions (Camille et al., 2011). Note, however, that the lesions of ACC patients in the study by Camille et al. (2011) were located in a more dorsal region of the ACC as compared to a more ventral peak activation in the present study. Interestingly, whereas increased activity of the ACC has been suggested for negative PEs (Holroyd and Coles, 2002), activity of the anterior ACC in the present study was positively correlated with PEs, i.e. outcomes better than expected led to enhanced activity of this ACC region. Rogers et al. (2004) reported increased BOLD signals in the subcallosal ACC following good outcomes, which may suggest a dissociation between different ACC regions in evaluating the valence of events.

In the present study, activations related to the action-independent PE also emerged bilaterally in the insula in both groups. For the action-dependent PE, however, the right insula was activated only in active learners, in which the modulation was also stronger as compared to observational learners, resembling the pattern for the dorsal striatum and the anterior cingulate (see above). Increased insular activity for erroneous action outcomes has previously been reported (Ullsperger et al., 2010). Interestingly, this finding is independent from agency (own vs. observed actions) when outcomes have low emotional impact as in correct vs. incorrect feedback (De Bruijn et al., 2009), but insular activity is stronger for own actions resulting in pain (Koban et al., 2013). This interaction is in line with the present findings of stronger modulations in active learners' insula activity for action-dependent monetary outcome PEs, which may be associated with stronger emotional responses as compared to correct vs. incorrect feedback.

Notably, a bias to significantly more right- than left-choices was present in the chance condition for active learners only. This bias may have resulted from the right-handedness of all subjects in the present study, as right-handers tend to associate the rightward space with positive concepts (Casasanto, 2009). This, in turn, may have triggered right-choices when left-choices did not differ in value, as was the case in the study by Casasanto (2009) and in the chance condition of the present study (for choice alternatives differing in value, as was the case for the 67% and 83% conditions, subjects did not show a right-bias but learned from feedback to make choices based on these values). Casasanto (2009, 2011) explained this bias by people's mental simulation of action execution, with right-handers' simulation of right-hand actions being more fluent, thereby leading to a preference of the rightward space, as this is usually affected by right-hand actions. Importantly, whereas active learners could mentally simulate their left or right button press prior to action execution, observational learners always had to press the same button in order to observe an active learner's choice in the learning blocks of the present study. Therefore, only active but not observational learners could experience more fluent mental simulations and, thus, a preference of right-choices in the learning blocks. In test blocks, where both groups could mentally simulate the execution of their own choices, the right-hand bias acquired in the learning trials may have been transferred by active learners, explaining the between-group difference in the bias. Interestingly, a numerical increase of right-choices from the first to the second test block is also seen in observers.

Considering that the bias to more right-choices was less pronounced in observational learners, it is necessary to discuss, whether this between-group difference in behaviour may account for the main finding of the present study, that is, between-group differences in neural processing of action-dependent but not action-independent PEs. If the bias towards more right- than left-choices in active learners corresponded to higher subjective values of the chance-condition cue and/or chance-condition right-choice, the deviation between model-based and real subjective values would be higher in active learners as compared to observers. Consequently, the learning model would have yielded a less accurate fit for active learners, which – assuming that active learners and observers did not differ regarding neural processing of PEs – would have led to a reduced detection of PE-related neural activity in active learners. Importantly, however, the opposite pattern, i.e. increased activity in ROIs related to processing of action-dependent PEs, was found in active learners relative to observers. It thus seems that the more fluent simulation of right-hand actions in the active learners (see above) provides an explanation of the right-bias, which is, however, unlikely to account for between-group differences in neural processing of PEs as reported in the present study.

4.2. Processing of PEs in the cerebellum and its connection with the PFC

The function of the cerebellum has classically been linked to motor control, and later studies (e.g. Gao et al., 1996; Inoue et al., 1998; Jueptner et al., 1997; for a review, see Blakemore and Sirigu, 2003) suggest that the cerebellum fulfils this function by performing calculations to predict the sensory consequences of actions in order to improve motor responses. In line with this view, the cerebellum was found to encode sensory PEs (Schlerf et al., 2012). Importantly, cerebellar neurons also code outcome PEs (Schultz and Dickinson, 2000), with activity changes in the vermis especially for unexpected rewards (Ramnani et al., 2004). Furthermore, a recent study shows increased activation of the lateral cerebellum in the outcome phase if the outcome was preceded by a cue of high predictive value (Lam et al., 2013). The present study adds to these findings by showing strong cerebellar activations both for action-independent and action-dependent outcome PEs, which, however, emerged only for active learners in separate analyses. Additionally, cerebellar activity was more strongly modulated by action-dependent PEs in active as compared to observational learners. This pattern is similar to stronger activations of the reward system for action-dependent PEs in active learners (see Section 4.1), indicating comparable PE coding in the reward system and the cerebellum. Interestingly, cerebellar PE coding may not depend on the reward system: whereas impaired learning from positive feedback in PD patients was suggested to result from reduced reward-related activity of the BG (Frank et al., 2004), cerebellar reward-related activity was still prominent in PD patients (Kunig et al., 2000). Accordingly, it is conceivable that outcome PE processing serving action selection is based on reward value more in the BG than in the cerebellum, which may be more generally involved in the prediction of events to support behavioural adaptation. This view is further supported by cerebellar loops with the PFC (Kelly and Strick, 2003), which is also involved in generating predictions (e.g. Alexander and Brown, 2011; Rogers et al., 2004).

In addition to processing of outcome PEs, activity of the cerebellum also depended on the cue-related uncertainty in observational learners, which is in line with uncertainty-related cerebellar activations in previous (Blackwood et al., 2004; Keri et al., 2004) studies.

Interestingly, uncertainty-related cerebellar activations were found only in observational learners. The explanation for this finding may consist of at least two aspects: first, cerebellar function has been associated with shifts of attention (Courchesne and Allen, 1997; Le et al., 1998). Second, uncertainty is regarded as the extent to which attention should be paid to the cue (Prevost et al., 2011). Accordingly, observational learners' attention may be more focused on the cue, presumably leading to increased modulation of cerebellar activity by uncertainty. Active learners, on the other hand, may focus more on cue-dependent action selection and preparation, which is not necessary in observational learning at the time of cue presentation.

In the present study, uncertainty-related activations were also found in a part of the right medial frontal cortex, which belongs to the PFC. The role of the PFC has been associated with decision-making under both risk and uncertainty (e.g. Bechara et al., 1999; Fellows and Farah, 2005; Hsu et al., 2005; Sanfey et al., 2003). The present study shows that especially uncertainty-related activations in the PFC are larger for active than observational learners. Taken together, the PFC may not process uncertainty per se but rather when it serves decision-making as necessary in active but not observational learners.

4.3. The processing of SVs and AVs in the reward system and frontal cortex

Action-independent outcome PEs involve updating SVs, that is, they increase and decrease with positive and negative outcomes, respectively. Importantly, the SV is independent from the action that led to a positive or negative outcome. Similar to the SV, the AV increases and decreases with positive and negative outcomes, respectively, whereas the value of the unselected action changes inversely (see Section 2.4.4 for details).

As AVs and SVs may be associated with processes of action preparation, outcome expectation, or both, we also explored in which brain regions activity was modulated by these parameters at the time of cue presentation. Furthermore, as outcome expectations may also or alternatively arise following an action, AV-related activity following (active or observed) choices was also analysed.

Notably, whereas previous studies have shown activity related to SVs in the striatum and anterior regions of the PFC (e.g. Kable and Glimcher, 2007; Peters and Buchel, 2009; Pine et al., 2009), the present study revealed significant modulations only in the medial frontal cortex. At least two factors may account for the absence of modulations in the striatum and anterior PFC regions. On the one hand, in order to include the cerebellum completely, EPIs were not tilted – presumably at the cost of sensitivity especially for OFC regions, which was shown to improve by application of a tilt (Deichmann et al., 2003). On the other hand, the absence of SV-related activity in the anterior PFC and the striatum may result from a 'distribution' of the reward value across two events, that is, presentation of a cue and a subsequent choice between different stimuli (arrows), with neither of these events alone indicating reward (see Section 4.1). Consequently, activity in the PFC and striatum may be more closely linked to the reward value associated with a combination of SVs and AVs than SVs alone, the latter of which indicate only *potential* rewards which require a further event such as the selection of an action. In line with this view, AV-related activations, which have previously been reported in parts of the reward system, such as in the PFC and the putamen (FitzGerald et al., 2012), emerged in the left putamen of active learners following responses but not cues in the present study. A possible explanation for the absence of this effect in observational learners refers to stronger striatal activations for links between own as compared to observed actions and expected outcomes (Bellebaum et al., 2012), which is also in line with the results of the present study on differences between active and observational learners' processing of action-dependent but not action-independent PEs (see Section 4.1).

5. Conclusions

The present study demonstrates that processing of outcome PEs in the striatum, the anterior cingulate, the insula, and the cerebellum is stronger in active than observational learning only for outcome PEs which depend on the preceding action, but not for those which depend on the initial cue. Consequently, processing of outcome PEs depends less on the learning type (active vs. observation) than the link to (own) actions. Importantly, learning of stimulus–action–outcome associations does not depend on action-independent but on action-dependent PEs, which is a prerequisite for (stimulus–)response–outcome learning both in active and observational learning. Despite stronger striatal and cerebellar modulations in active learners, learning performance was comparable in observational learners and was most likely achieved by recruitment of brain structures beyond the reward system (Monfardini et al., 2013). Furthermore, learning strategies regarding e.g. memorising stimulus–response–outcome associations may differ between groups. Both of these aspects need to be addressed further in future research.

The present study did not yield a clear picture concerning the role of the medial temporal lobe in active and observational learning from feedback. In particular, we did not find regions being more strongly involved in action-dependent PE coding in observational than active learning. This question requires further investigation in the future.

Role of the funding source

This research was funded by the Ministry of Innovation, Science and Research of the Federal State of Nordrhein-Westfalen, Germany (Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen; http://www.wissenschaft.nrw.de; Grant number 334-4). The funder had no role in study design, data collection, analysis, and interpretation, decision to publish, or writing of the report.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2014.10.036.

References

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action–outcome predictor. Nat. Neurosci. 14, 1338–1344.

Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modeling geometric deformations in EPI time series. Neuroimage 13, 903–919.

Bechara, A., Damasio, H., Damasio, A.R., Lee, G.P., 1999. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. J. Neurosci. 19, 5473–5481.

Bellebaum, C., Jokisch, D., Gizewski, E.R., Forsting, M., Daum, I., 2012. The neural coding of expected and unexpected monetary performance outcomes: dissociations between active and observational learning. Behav. Brain Res. 227, 241–251.

Bellebaum, C., Kobza, S., Thiele, S., Daum, I., 2010. It was not my fault: event-related brain potentials in active and observational learning from feedback. Cereb. Cortex 20, 2874–2883.

Bellebaum, C., Koch, B., Schwarz, M., Daum, I., 2008. Focal basal ganglia lesions are associated with impairments in reward-based reversal learning. Brain 131, 829–841.

Blackwood, N., Ffytche, D., Simmons, A., Bentall, R., Murray, R., Howard, R., 2004. The cerebellum and decision making under uncertainty. Brain Res. Cogn. Brain Res. 20, 46–53.

Blakemore, S.J., Sirigu, A., 2003. Action prediction in the cerebellum and in the parietal lobe. Exp. Brain Res. 153, 239–245.

Buchel, C., Morris, J., Dolan, R.J., Friston, K.J., 1998. Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron 20, 947–957.

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

Camille, N., Tsuchida, A., Fellows, L.K., 2011. Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. J. Neurosci. 31, 15048–15052.

Casasanto, D., 2009. Embodiment of abstract concepts: good and bad in right- and left-handers. J. Exp. Psychol. – Gen. 138, 351–367.

Casasanto, D., Chrysikou, E.G., 2011. When left is 'Right': motor fluency shapes abstract concepts. Psychol. Sci. 22, 419–422.

Casey, B.J., Trainor, R., Giedd, J., Vauss, Y., Vaituzis, C.K., Hamburger, S., Kozuch, P., Rapoport, J.L., 1997. The role of the anterior cingulate in automatic and controlled processes: a developmental neuroanatomical study. Dev. Psychobiol. 30, 61–69.

Clark, L., Lawrence, A.J., Astley-Jones, F., Gray, N., 2009. Gambling near-misses enhance motivation to gamble and recruit win-related brain circuitry. Neuron 61, 481–490.

Courchesne, E., Allen, G., 1997. Prediction and preparation, fundamental functions of the cerebellum. Learn. Mem. 4, 1–35.

Daum, I., Schugens, M.M., Ackermann, H., Lutzenberger, W., Dichgans, J., Birbaumer, N., 1993. Classical-conditioning after cerebellar lesions in humans. Behav. Neurosci. 107, 748–756.

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

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441.

Delgado, M.R., 2007. Reward-related responses in the human striatum. Ann. N. Y. Acad. Sci 1104, 70–88.

Delgado, M.R., Miller, M.M., Inati, S., Phelps, E.A., 2005. An fMRI study of reward-related probability learning. Neuroimage 24, 862–873.

Delgado, M.R., Nystrom, L.E., Fissell, C., Noll, D.C., Fiez, J.A., 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. J. Neurophysiol. 84, 3072–3077.

Dickerson, K.C., Li, J.A., Delgado, M.R., 2011. Parallel contributions of distinct human memory systems during probabilistic learning. Neuroimage 55, 266–276.

Fellows, L.K., Farah, M.J., 2005. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. Cereb. Cortex 15, 58–63.

FitzGerald, T.H.B., Friston, K.J., Dolan, R.J., 2012. Action-specific value signals in reward-related regions of the human brain. J. Neurosci. 32, 16417–16423.

FitzGerald, T.H.B., Seymour, B., Dolan, R.J., 2009. The role of human orbitofrontal cortex in value comparison for incommensurable objects. J. Neurosci. 29, 8388–8395.

Frank, M.J., Seeberger, L.C., O'reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940–1943.

Friston, K.J., Glaser, D.E., Henson, R.N.A., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. Neuroimage 16, 484–512.

Fudenberg, D., Levine, D., 1998. Learning in games. Eur. Econ. Rev. 42, 631–639.

Gao, J.H., Parsons, L.M., Bower, J.M., Xiong, J.H., Li, J.Q., Fox, P.T., 1996. Cerebellum implicated in sensory acquisition and discrimination rather than motor control. Science 272, 545–547.

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

Garavan, H., Ross, T.J., Murphy, K., Roche, R.A.P., Stein, E.A., 2002. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. Neuroimage 17, 1820–1829.

Glascher, 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. Cereb. Cortex 19, 483–495.

Gluck, M.A., Bower, G.H., 1988. From conditioning to category learning – an adaptive network model. J. Exp. Psychol. Gen. 117, 227–247.

Guitart-Masip, M., Chowdhury, R., Sharot, T., Dayan, P., Duzel, E., Dolan, R.J., 2012. Action controls dopaminergic enhancement of reward representations. Proc. Natl. Acad. Sci. USA 109, 7511–7516.

Haber, S.N., Fudge, J.L., 1997. The primate substantia nigra and VTA: integrative circuitry and function. Crit. Rev. Neurobiol. 11, 323–342.

Hester, R., Barre, N., Murphy, K., Silk, T.J., Mattingley, J.B., 2008. Human medial frontal cortex activity predicts learning from errors. Cereb. Cortex 18, 1933–1940.

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

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R.B., Coles, M.G., Cohen, J.D., 2004. Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. Nat. Neurosci. 7, 497–498.

Hoshi, E., Tremblay, L., Feger, J., Carras, P.L., Strick, P.L., 2005. The cerebellum communicates with the basal ganglia. Nat. Neurosci. 8, 1491–1493.

Houk, J.C., Adams, J.L., Barto, A.G., 1995. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, J.C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, pp. 249–270.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C.F., 2005. Neural systems responding to degrees of uncertainty in human decision-making. Science 310, 1680–1683.

Inoue, K., Kawashima, R., Satoh, K., Kinomura, S., Goto, R., Koyama, M., Sugiura, M., Ito, M., Fukuda, H., 1998. PET study of pointing with visual feedback of moving hands. J. Neurophysiol. 79, 117–125.

Joel, D., Niv, Y., Ruppin, E., 2002. Actor–critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. 15, 535–547.

Jueptner, M., Ottinger, S., Fellows, S.J., Adamschewski, J., Flerich, L., Muller, S.P., Diener, H.C., Thilmann, A.F., Weiller, C., 1997. The relevance of sensory input for the cerebellar control of movements. Neuroimage 5, 41–48.

Kable, J.W., Glimcher, P.W., 2007. The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. 10, 1625–1633.

Kelly, R.M., Strick, P.L., 2003. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. J. Neurosci. 23, 8432–8444.

Keri, S., Decety, J., Roland, P.E., Gulyas, B., 2004. Feature uncertainty activates anterior cingulate cortex. Hum. Brain Mapp. 21, 26–33.

Knutson, B., Cooper, J.C., 2005. Functional magnetic resonance imaging of reward prediction. Curr. Opin. Neurol. 18, 411–417.

Koban, L., Corradi-Dell'Acqua, C., Vuilleumier, P., 2013. Integration of error agency and representation of others' pain in the anterior insula. J. Cogn. Neurosci. 25, 258–272.

Kobza, S., Ferrea, S., Schnitzler, A., Pollok, B., Sudmeyer, M., Bellebaum, C., 2012. Dissociation between active and observational learning from positive and negative feedback in Parkinsonism. PLoS One 7, e50250.

Kunig, G., Leenders, K.L., Martin-Solch, C., Missimer, J., Magyar, S., Schultz, W., 2000. Reduced reward processing in the brains of Parkinsonian patients. Neuroreport 11, 3681–3687.

Labar, K.S., Gatenby, J.C., Gore, J.C., Ledoux, J.E., Phelps, E.A., 1998. Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron 20, 937–945.

Lam, J.M., Wachter, T., Globas, C., Karnath, H.O., Luft, A.R., 2013. Predictive value and reward in implicit classification learning. Hum. Brain Mapp. 34, 176–185.

Le, T.H., Pardo, J.V., Hu, X.P., 1998. 4T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. J. Neurophysiol. 79, 1535–1548.

Lehrl, S., 2005. Manual Zum MWT-B, Balingen. Spitta-Verlag.

Lieberman, M.D., Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc. Cogn. Affect. Neurosci. 4, 423–428.

Liljeholm, M., Molloy, C.J., O'Doherty, J.P., 2012. Dissociable brain systems mediate vicarious learning of stimulus–response and action–outcome contingencies. J. Neurosci. 32, 9878–9886.

Menard, S., 1995. Applied Logistic Regression Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage, Thousand Oaks, pp. 07–106.

Monfardini, E., Gazzola, V., Boussaoud, D., Brovelli, A., Keysers, C., Wicker, B., 2013. Vicarious neural processing of outcomes during observational learning. PLoS One 8, e73879.

Myers, R., 1990. Classical and Modern Regression with Applications, 2nd ed. Duxbury, Boston.

Myung, I.J., Busemeyer, J.R., 1992. Measurement-free tests of a general state-space model of prototype learning. J. Math. Psychol. 36, 32–67.

Niv, Y., Daw, N.D., Joel, D., Dayan, P., 2007. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191, 507–520.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J., 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304, 452–454.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C., 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat. Neurosci. 4, 95–102.

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S., 2002. Activity in human ventral striatum locked to errors of reward prediction. Nat. Neurosci. 5, 97–98.

Penny, W.D., Holmes, A.D., 2003. Random effects analysis. In: Frackowiak, R.S.J., Friston, K., Frith, C.D., Dolan, R., Price, C.J., Zeki, S.A.J., Ashburner, J.T., Penny, W.D. (Eds.), Human Brain Function. Academic Press, San Diego, pp. 843–850.

Peters, J., Buchel, C., 2009. Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. J. Neurosci. 29, 15727–15734.

Pine, A., Seymour, B., Roiser, J.P., Bossaerts, P., Friston, K.J., Curran, H.V., Dolan, R.J., 2009. Encoding of marginal utility across time in the human brain. J. Neurosci. 29, 9575–9581.

Prevost, C., Mccabe, J.A., Jessup, R.K., Bossaerts, P., O'Doherty, J.P., 2011. Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. Eur. J. Neurosci. 34, 134–145.

Ramnani, N., Elliott, R., Athwal, B.S., Passingham, R.E., 2004. Prediction error for free monetary reward in the human prefrontal cortex. Neuroimage 23, 777–786.

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

Rogers, R.D., Ramnani, N., Mackay, C., Wilson, J.L., Jezzard, P., Carter, C.S., Smith, S.M., 2004. Distinct portions of anterior cingulate cortex and medial prefrontal cortex are activated by reward processing in separable phases of decision-making cognition. Biol. Psychiatry 55, 594–602.

Rolls, E.T., McCabe, C., Redoute, J., 2008. Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. Cereb. Cortex 18, 652–663.

Sanfey, A.G., Hastie, R., Colvin, M.K., Grafman, J., 2003. Phineas gauged: decision-making and the human prefrontal cortex. Neuropsychologia 41, 1218–1229.

Schlerf, J., Ivry, R.B., Diedrichsen, J., 2012. Encoding of sensory prediction errors in the human cerebellum. J. Neurosci. 32, 4913–4922.

Schultz, W., 1997. Dopamine neurons and their role in reward mechanisms. Curr. Opin. Neurobiol. 7, 191–197.

Schultz, W., 1998a. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27.

Schultz, W., 1998b. The phasic reward signal of primate dopamine neurons. Adv. Pharmacol. 42, 686–690.

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

Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500.

Shiner, T., Seymour, B., Wunderlich, K., Hill, C., Bhatia, K.P., Dayan, P., Dolan, R.J., 2012. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. Brain 135, 1871–1883.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart, J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Brain Res. Cogn. Brain Res. 17, 75–82.

Smittenaar, P., Chase, H.W., Aarts, E., Nusselein, B., Bloem, B.R., Cools, R., 2012. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection <endash> learning or performance. Eur. J. Neurosci. 35, 1144–1151.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, New York.

Thoma, P., Bellebaum, C., Koch, B., Schwarz, M., Daum, I., 2008. The cerebellum is involved in reward-based reversal learning. Cerebellum 7, 433–443.

Ullsperger, M., Harsay, H.A., Wessel, J.R., Ridderinkhof, K.R., 2010. Conscious perception of errors and its relation to the anterior insula. Brain Struct. Funct. 214, 629–643.

Ullsperger, M., 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.

van der Helden, J., Boksem, M.A., Blom, J.H., 2010. The importance of failure: feedback-related negativity predicts motor learning efficiency. Cereb. Cortex 20, 1596–1603.

Wunderlich, K., Rangel, A., O'Doherty, J.P., 2009. Neural computations underlying action-based decision making in the human brain. Proc. Natl. Acad. Sci. USA 106, 17199–17204.

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

Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., McGill, K., Jaggi, J.L., Baltuch, G.H., Kahana, M.J., 2009. Human substantia nigra neurons encode unexpected financial rewards. Science 323, 1496–1499.