Neuroscience and Biobehavioral Reviews 37 (2013) 1297–1310

Contents lists available at SciVerse ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review

Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies

Jane Garrison[a,1], Burak Erdeniz[a,b], John Done[a,*]

[a] School of Psychology, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK
[b] School of Computer Science, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form 19 March 2013
Accepted 27 March 2013

Keywords:
Prediction error
Reward
Punishment
Striatum
Habenula
Actor–critic
Reinforcement learning

ABSTRACT

Activation likelihood estimation (ALE) meta-analyses were used to examine the neural correlates of prediction error in reinforcement learning. The findings are interpreted in the light of current computational models of learning and action selection. In this context, particular consideration is given to the comparison of activation patterns from studies using instrumental and Pavlovian conditioning, and where reinforcement involved rewarding or punishing feedback. The striatum was the key brain area encoding for prediction error, with activity encompassing dorsal and ventral regions for instrumental and Pavlovian reinforcement alike, a finding which challenges the functional separation of the striatum into a dorsal 'actor' and a ventral 'critic'. Prediction error activity was further observed in diverse areas of predominantly anterior cerebral cortex including medial prefrontal cortex and anterior cingulate cortex. Distinct patterns of prediction error activity were found for studies using rewarding and aversive reinforcers; reward prediction errors were observed primarily in the striatum while aversive prediction errors were found more widely including insula and habenula.

© 2013 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author. Tel.: +44 1707 284638; fax: +44 1707 285073.
E-mail address: d.j.done@herts.ac.uk (J. Done).
[1] Present address: Department of Psychology, and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

0149-7634/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neubiorev.2013.03.023

1. Introduction

Decision making requires learning of associations between conditioned stimuli and outcomes (Pavlovian learning), or between one's actions and their consequences (instrumental learning), that are rewarding or punishing. The behavioural literature on reinforcement learning has demonstrated that it is not the reward (or punishment) per se that reinforces (extinguishes) behaviours but the difference between the predicted value of future rewards (punishments) and their realised value. This is known as the reward prediction error (RPE). In the original quantitative models of reinforcement learning (Bush and Mosteller, 1951; Mackintosh, 1975; Pearce and Hall, 1980; Rescorla and Wagner, 1972) the effect of unexpected outcomes on reinforcement learning is calculated as the difference between the reward received and reward expected. This is known as the prediction error (PE) and is depicted by the following formula:

$$\delta = R_t - V_{(t)} \quad (1)$$

in which R_t is the value of reward received (or unconditioned stimulus value, US) and V_t is the expected value of reward signified by the conditioned stimulus (CS), both at time t. When this prediction error (PE) equates to zero then learning does not occur, even when there continues to be a joint occurrence of conditioned and unconditioned stimuli (Schultz and Dickinson, 2000; Niv and Schoenbaum, 2008). Sutton and colleagues noted various limitations inherent in the original PE model of Rescorla and Wagner and developed the temporal difference prediction method of learning (see Sutton and Barto, 1987), now known as the temporal difference learning algorithm (TD). In TD, prediction error becomes the difference between the expected value of all future reward at a certain point in time (deemed the state at time t) and the expected value of all future reward at the succeeding state (at time $t+1$). The TD prediction error is calculated using the following equation:

$$\delta_t = r_t + \gamma V(S_{t+1}) - V(S_t) \quad (2)$$

where r_t is the reward received, γ is the discount factor determining the weight given to future state values, and $V(S_t)$ and $V(S_{t+1})$ represent the value of the current state and the subsequent state respectively.

This simple formulation of TD prediction error can also be extended to incorporate situations involving action learning, most notably using the Q-learning and SARSA ('state, action, reward, state, and action') algorithms. In these models, action-value coding is undertaken by learning the reward value of state-action pairs, with action-value then updated using the prediction error (Kaelbling et al., 1996; Sutton and Barto, 1998; Dayan and Abbott, 2001). Q-learning and SARSA differ in this process: Q-learning uses an off-policy method whereby the estimated value function is updated using hypothetical actions whereas SARSA assumes an on-policy algorithm where the value function update is based strictly on experience (Sutton and Barto, 1998). In comparison to Q-learning and SARSA, the TD learning algorithm directly updates the reward value of states (rather than of state-action pairs) based on discrete periods of time between the CS and US. Therefore, it is more commonly used in Pavlovian conditioning experiments (i.e., O'Doherty et al., 2003). However it is important to note that the TD prediction error in Eq. (2) can also be modified to cope with instrumental-conditioning scenarios such as actor–critic and advantage learning (O'Doherty et al., 2004).

Given the relevance of prediction error to models of reinforcement learning, it is of no surprise that a large number of electrophysiology studies with animals and fMRI studies with humans have examined the brain regions involved in the computation of prediction errors. Furthermore, the electrophysiology studies have reported remarkably high similarity between reward prediction error (RPE) and the spiking activity of dopamine (DA) neurons in the midbrain (Montague et al., 1996; Schultz et al., 1997). There is thus additional interest in the study of RPE in humans using fMRI, since RPE may be taken as a proxy measure of DA related activity in both the midbrain and in areas such as the striatum, to which midbrain DA neurons project.

Computational models developed in machine reinforcement learning have been widely applied in the neuroscience literature and this is especially true for the human reinforcement learning literature where the regions of interest have involved the basal ganglia, and the neural circuitry that links this subcortical region to other subcortical and cortical brain regions. The most widely debated family of computational models come under the heading of the actor–critic framework (Houk et al., 1995; Suri and Schultz, 1998, 1999; Joel et al., 2002; Khamassi, 2005). There are various versions of the actor–critic model, but the general format comprises an actor module, which learns to select actions in order to maximise future reward together with a critic module, which calculates a TD prediction error (Barto, 1995). For the critic to calculate a prediction error, it must meet several criteria (Glimcher, 2011). More specifically, for the critic to operate as a TD calculator then a large TD error should result not just from an unexpected unconditioned stimulus, but also from an unexpected conditioned stimulus, so long as this is strongly associated with future gain or loss. The sign of the TD error will also vary according to whether the predicted, or received, reward is positive or negative.

Recent narrative reviews of human fMRI studies have concluded that RPE in both Pavlovian and instrumental learning is computed in the ventral striatum (VS), with the dorsal striatum (DS) involved only in instrumental learning (Lohrenz et al., 2007; O'Doherty et al., 2004; Porcelli and Delgado, 2009). If so, then this division of labour maps on well to the actor–critic model of reinforcement learning with the VS operating as the critic and the DS operating as the actor.

However, this neat structural mapping of actor and critic has its detractors. Firstly, it makes the assumption of a single critic whereas the true situation may be more complicated with the critic possibly partitioned functionally and structurally, having evolved to deal with different task requirements. From a biologically informed perspective, Redgrave and Gurney (2006) have argued that the critic may not be localised in one brain region but distributed.

Secondly, unlike machines, animals can accomplish a task via different routes through to the final action, referred to as 'model-based' (goal directed) or 'model-free' (automatic) processes (see Daw et al., 2005; Dayan, 2009 for a discussion). The actor–critic algorithm supports the model-free approach only (Balleine et al., 2008) whereas the model-based approach makes assumptions that require the existence of internally generated state transitions, including those that can be expected when preparing action selection (van der Meer and Redish, 2010).

The narrative reviews of fMRI studies of RPE referred to above, have been based on a small and select number of available studies rather than a systematic literature search followed by a meta-analysis. A recent meta-analysis that used activation likelihood estimation (ALE) to examine the neural correlates of rewards and punishments in 142 reward processing studies reported a widespread network of brain activations involving various prefrontal regions, striatum, inferior parietal lobe and insula (Liu et al., 2010). This review did not select specifically for reinforcement learning studies (it included a diverse range of decision making tasks) and utilised objective reward/punishment values, rather than prediction error. Furthermore Liu et al. (2010) did not disaggregate Pavlovian and instrumental forms of learning, and hence their findings could not contribute to the debate about the validity of computational models of reinforcement learning as useful models of the functional organisation of human striatum.

fMRI studies of prediction error are typically based on a Rescorla–Wagner or TD modelled prediction error implemented in various learning algorithms (e.g., advantage learning, Q-learning, SARSA) in which the estimated PE is calculated for each stimulus event. This time series is then regressed onto the series of fMRI images to identify those voxels in which the BOLD activation value correlates with estimated PE (O'Doherty et al., 2007). The issue of valency of outcome (reward vs. punishment, monetary gain vs. loss) remains controversial with some suggestions of a possible separation of brain systems calculating PE for rewards and punishments respectively, e.g. reward PEs are calculated in the striatum and punishment PEs are calculated in various cortical regions including anterior cingulate cortex (ACC) and insula (Nieuwenhuis et al., 2005). In fact, the Liu et al. (2010) meta-analysis found a high degree of overlap in the brain regions coding both expected or experienced gains and losses.

In the following meta-analysis we examine RPE and aversive prediction error (APE) in reinforcement learning paradigms involving Pavlovian and instrumental conditioning. To our knowledge there have been no previous meta-analytic reviews of the numerous fMRI studies based on the parametric modelling of prediction errors. We structured our meta-analysis around the following research questions: Is reward prediction error processing widespread or principally computed within ventral and dorsal striatum? Do differential patterns of activation for Pavlovian and instrumental prediction errors implicate an actor–critic organisation in the basal ganglia? And finally, to what extent do activation patterns overlap or segregate for reward and punishment PEs?

2. Methods

2.1. Systematic literature search

Studies were selected for the meta-analysis by searching the SciVerse Scopus (http://www.scopus.com/) and Pubmed (http://www.pubmed.org/) databases using the following search terms: "fmri OR neuroimaging" AND "prediction error" AND ("reinforcement learning" OR "classical conditioning" OR Pavlovian OR instrumental OR reward). The BrainMap Sleuth database (http://brainmap.org) was searched using the "prediction error" term, and reference lists from relevant review articles were assessed together with lists of articles written by key researchers in the field. These search results were merged, with duplicates eliminated, to yield a total of 779 articles (October 2011). This high number is explained by the diverse terminology within this topic area and its high level of overlap with similar, but distinct, fields of study. Attempts to reduce the size of this initial screening by removing search terms eliminated studies of potential interest; the initial screen was therefore maintained at this broad level to minimise the risk of missing relevant articles.

Abstracts from the above 779 articles were reviewed. A large number of studies were rejected which either made reference to prediction error without being primarily focused on reinforcement learning (primary reason for discard), or which failed to provide coordinates for areas of relevant brain activation. An intermediate shortlist of 109 articles of interest was then constructed, which were studied in detail. Inclusion and exclusion criteria were applied to these papers. Inclusion criteria were as follows: (1) primary research studies using human adult participants; (2) prediction error calculated using Rescorla Wagner or TD models, or from models mathematically similar to these; (3) coordinates of prediction error provided in Montreal Neurological Institute (MNI) or Talairach standard stereotactic space; (4) the study involved use of an experimental reinforcement learning task providing subject feedback (n.b. studies were excluded where tasks involved the simple probabilistic allocation of reward or punishment such as monetary incentive delay tasks). Studies were also excluded from the meta-analysis of whole brain activation studies if (1) analysis was based solely on one or more regions of interest (ROIs) e.g. using anatomical masks or based on coordinates from other studies; and (2) where sample populations were investigated whose brain functions might be expected to deviate from those of normal healthy adults (e.g. aged population, Parkinson's disease patients, substance dependent adults) although separately reported results for a matched control group were included if coordinate data was available.

Email contact was made with the authors of 14 papers where the study met the inclusion criteria but where whole brain prediction error coordinates were not included in the original articles. We had replies from 12 of which an additional 5 provided data (see Acknowledgements) making a total of 35 whole-brain activation studies. These are listed in Appendix A.

Initial screening of studies and categorisation was undertaken by the first author. Papers where there was uncertainty over selection were retained for detailed analysis and discussion by the co-authors until agreement was reached. These papers were re-read by at least two of the authors to agree the specific issues that would affect analysis, namely the experimental method, the model used for measuring PE, the criteria used for voxel-level thresholding (whether whole brain analyses, ROI or if small volume corrected (SVC) data). In 32 of the 35 studies computational models similar to the Rescorla–Wagner or TD method were used to calculate prediction errors (whether positive or negative) on a trial-by-trial basis. These models also permitted the use of regressors for estimates of action values (Q-learning, SARSA) or state values of the CS in addition to prediction error. Whereas prediction error was typically time-stamped to the onset of the US, action or state values were time stamped to the CS. The other 3 studies (McClure et al., 2003; Morris et al., 2011; Ramnani et al., 2004) did not adopt computational models but nevertheless categorised expected and unexpected outcomes and conducted subtraction analyses to estimate a prediction error. The majority of studies (54%) used an uncorrected threshold of typically $p < .001$ (one used $p < .0001$, another $p < .005$) at the voxel level with a cluster size typically within the range 5–20 contiguous voxels, although this was not always described. A further 37% of studies used more conservative voxel level thresholds (typically $p < .05$ false discovery rate (FDR) or family-wise error (FWE) corrected).

Although we excluded all papers which used a regions of interest (ROI) analysis, one study (Valentin et al., 2009) used a $p < .001$ uncorrected threshold for striatum with a $p < .05$ FWE threshold for all other brain areas. The same thresholding method was used in three further studies (Hampton et al., 2006; Seymour et al., 2005, 2007). However, for these studies the regions of interest specified in the a priori hypothesis were extensive, encompassing from six to eleven separate brain regions. Since these a priori hypotheses adopted an acceptable strategy that should not compromise the results they were included in the analysis. Indeed exclusion of these studies did not make any material difference to the results

2.2. Study categorisation and extraction of coordinate data

Studies were categorised according to whether the experimental task involved instrumental or Pavlovian conditioning, with reward, punishment or a combination of both. MNI or Talairach based (x, y, z) coordinates for the foci of areas of BOLD activity associated with RPE or APE were extracted from each of the fMRI studies, with those coordinates listed in Talairach space converted to MNI using the icmb2tal algorithm implemented in the BrainMap's GingerALE 2.1 software (http://www.brainmap.org/ale/). All RPE and APE prediction error foci represented by both positive and

negative BOLD signals were extracted from each of the study papers and used in the meta-analyses. A master-list of all studies was created by combining all coordinates in MNI space in preparation for the ALE meta-analyses, with a total of 446 foci identified across the 35 studies.

Studies were grouped by experimental task to enable contrasts to be undertaken between the different task-types (instrumental vs. Pavlovian or reward vs. punishment). The minimum number of foci allocated to a group (Punishment) was 71, exceeding the minimum number of coordinates recommended for use of the GingerALE meta-analysis for a simple sensory task of 20 (http://www.brainmap.org/ale/). A summary of the studies used and their allocation to each meta-analysis group is shown in Table 1, with the studies listed in Appendix A.

In our analysis, dorsal and ventral striatum are given particular consideration due to their prominence in the reinforcement learning and decision making literatures. However no unequivocal boundary can be agreed (Voorn et al., 2004). It is now common for VS to be specified to include nucleus accumbens (NAcc), the olfactory tubercle and ventral portions of caudate and putamen (Haber and McFarland, 1999; Martin, 2003; Joel et al., 2002). Assignations of DS and VS in this paper have followed the broad definitions set out in Postuma and Dagher (2006), i.e.:

Putamen	DS :	$z > +2$	VS :	$z < +2$
Caudate	DS :	$z > +7$	VS :	$z < +7$

To assess potential NAcc activation, reported coordinates of VS prediction error activation were reviewed against the algorithm for recognition of the NAcc, and the MNI extent data as specified in Ahsan et al. (2007):

$$\text{Left hemisphere}: \quad -15 < x < -3; \quad 4 < y < 14; \quad -14 < z < -3$$
$$\text{Right hemisphere}: \quad 14 < x < 4; \quad 5 < y < 15; \quad -14 < z < -3$$

Foci were assigned to the NAcc if they fall within these extremes, which, by specifying a cuboid space, intentionally allows for a generous interpretation of NAcc activity. Activation coordinates are assigned as left or right in accordance with general neurological and MNI convention whereby a positive value of the x coordinate indicates a location in the right brain hemisphere.

2.3. Activation likelihood estimation (ALE) meta-analysis

ALE (Eickhoff et al., 2009; Laird et al., 2005; Turkeltaub et al., 2002) is a coordinate based quantitative meta-analysis method that identifies consistent brain activation locations elicited across studies employing similar experimental conditions. In a comparison of alternative coordinate-based meta-analytical approaches (ALE, kernel density analysis and multi-level kernel density analysis), ALE was found to produce results most comparable to image based meta-analysis (IBMA; Salimi-Khorshidi et al., 2009). Without access to the full image data for each study required for IBMA, ALE is a

Table 1
Categorisation of fMRI prediction error studies and allocation to meta-analysis contrast groups.

Study	Type of prediction error	Number of subjects	All Studies	Inst	Pav	Reward	Punish	Combined R&P
Brovelli et al. (2008)	RW	14	√	√				√
Cohen et al. (2007)	RW	17	√	√		√		
Delgado et al. (2008)	TD	11	√		√		√	
Gläscher et al. (2010)	SARSA	18	√	√		√		
Gläscher et al. (2009)	RW	20	√	√				√
Gradin et al. (2011)	SARSA	17	√	√		√		
Hampton et al. (2006)	RW	16	√	√				√
Howard-Jones et al. (2010)	RW	16	√	√		√		
Jocham et al. (2011)	RW	16	√	√		√		
Kahnt et al. (2011)	RW	20	√	√				√
Kahnt et al. (2009)	RW	19	√	√		√		
Kim et al. (2006)	TD	16	√	√		√	√	
Kumar et al. (2008)	TD	18	√		√	√		
Landmann et al. (2007)	RW	16	√	√		√		
Li et al. (2011a)	RW	20	√	√				√
Li et al. (2006)	RW	46	√	√				√
Li et al. (2011b)	RW	17	√		√		√	
McClure et al. (2003)	N/A	28	√		√	√		
Morris et al. (2011)	N/A	16	√		√	√		
Murray et al. (2008)	RW	12	√	√		√		
O'Doherty et al. (2003)	TD	13	√		√	√		
O'Sullivan et al. (2011)	RW	24	√	√		√		
Park et al. (2010)	RW	16	√	√				√
Pessiglione et al. (2006)	RW	13	√	√		√		
Ramnani et al. (2004)	N/A	6	√		√	√		
Rodriguez et al. (2006)	RW	15	√	√				√
Schonberg et al. (2010)	TD	17	√	√				√
Seger et al. (2010)	SARSA	10	√	√				√
Seymour et al. (2007)	TD	24	√		√	√	√	
Seymour et al. (2005)	TD	19	√		√	√	√	
Spoormaker et al. (2011a)	TD	35	√		√		√	
Spoormaker et al. (2011b)	TD	18	√		√		√	
Tanaka et al. (2006)	SARSA	18	√	√				√
Tobler et al. (2006)	TD	22	√		√	√		
Valentin et al. (2009)	RW	17	√	√		√		
Number of studies			35	23	12	20	7	11
Number of foci			446	293	153	262	71	120

For simplicity, meta-analytical studies are referred to by the name of the first author and year of publication. Inst: instrumental; Pav: Pavlovian. RW: Rescorla Wagner; TD: temporal-difference; SARSA: State-Action-Reward-State-Action. N/A = no explicit model fitting applied, but the contrasts analysed in the second level fit with the basic assumptions of RW prediction error. The algorithms that use TD used either the actor–critic or advantage learning rules.

Table 2
Detailed ALE cluster results for all prediction error studies.

All studies	L/R	*x*	*y*	*z*	ALE (10^{-3})	*Size*	Comments
Pallidum	L	−10	6	−4	54	10, 904	Cluster encompasses NAcc
Putamen (VS)	L	−18	6	−12	40		
Caudate (VS)	L	−12	6	6	32		
Putamen (DS)	L	−22	16	4	13		
Caudate (VS)	R	10	6	−2	60	8.768	Cluster encompasses NAcc
Putamen (DS)	R	22	4	12	20		
Claustrum	L	−32	22	−4	19	896	
Insula	L	−30	22	−10	19		
Anterior cingulate	R	10	50	−4	18	792	
Anterior cingulate	R	10	46	−6	18		
Medial frontal gyrus		0	28	42	23	792	
Cingulate gyrus	R	2	34	20	22	744	
Claustrum	R	34	22	−8	20	520	
Precentral gyrus	L	−50	6	30	17	408	
Inferior frontal gyrus	R	54	12	18	17	312	
Superior frontal gyrus	R	4	10	50	17	312	
Cingulate gyrus	L	−8	32	30	16	304	
Middle frontal gyrus	L	−42	42	−12	17	280	
Anterior cingulate		0	36	0	16	168	
Middle frontal gyrus	L	−32	54	0	16	168	
Red nucleus	L	−10	−22	−6	14	144	
Thalamus	L	−8	−26	−4	14		
Precentral gyrus	L	−40	−10	44	16	136	
Inferior parietal lobule	R	46	−40	54	15	112	
Inferior parietal lobule	L	−40	−42	50	15	96	
Middle temporal gyrus	R	58	−26	−18	15	88	
Posterior cingulate	R	4	−32	24	14	80	
Medial frontal gyrus	R	26	38	28	14	72	
Supramarginal gyrus	R	48	−50	36	16	72	
Cerebellum	L	−16	−46	−12	14	64	
Superior temporal gyrus	R	56	4	−8	14	64	
Middle frontal gyrus	L	−36	26	40	14	64	
Superior occipital gyrus	R	38	−82	34	18	56	

Abbreviations: DS, dorsal striatum; VS, ventral striatum; NAcc, nucleus accumbens; L, left hemisphere; R, right hemisphere.

preferred method for meta-analytical comparison of neuroimaging data.

In ALE, all foci reported in a given study are modelled by creating 3D Gaussian probability distributions centred at each foci (the reported *x*, *y*, *z* coordinates). In BrainMap's GingerALE 2.1 the width of the distribution, reflecting spatial uncertainty, is adjusted to accommodate between subject variance. The modelled probability distributions for all reported foci are then combined to form a modelled activation (MA) map for that experimental task for each study. Given the adjustment for sample size, studies with larger numbers of subjects have tighter Gaussian distributions for all foci within an MA and hence provide greater weight to those foci when combined in the meta-analysis. Following the union of MAs across studies, activation probabilities, or ALE scores, are determined for each voxel. To enable statistical inference about spatial patterns of activation, the null hypothesis assumes that spatial patterns of activation are associated randomly across studies. A null distribution is achieved by randomly sampling a voxel from one MA map, then repeating this for every other MA map and obtaining the union of activation probabilities in exactly the same way as for the real MAs. This process is repeated 10^{11} times to allow an ALE null distribution to be estimated against which the derived data may be assessed (Eickhoff et al., 2009).

For contrasts requiring subtraction between two experimental tasks (e.g. instrumental – Pavlovian to find which areas are more active for instrumental than for Pavlovian) the difference between ALE scores are calculated. A null distribution for this difference is then determined by pooling the data from both experimental tasks and randomly sampling from this pool to simulate two samples with similar numbers of foci as reported in the real data. Thus unequal numbers of foci between two experimental tasks are incorporated in the modelled null distribution for each subtraction analysis (Eickhoff et al., 2009). The nonparametric p values for the ALE maps for each experimental task are then thresholded using the false discovery rate (FDR) method. For this study the FDR was set at $p < .05$ with a minimum cluster volume of 50 mm^3 using 'all extrema' peak cluster analysis to aid identification of individual areas of activation within large single clusters. For the subtraction analyses, an uncorrected p value of .05 and a minimum cluster volume of 50 mm^3 were used. The use of the uncorrected *p*-value was adopted to avoid overly conservative results given that the inputs into these subtraction analyses have already been thresholded using FDR (Eickhoff et al., 2011).

Final ALE cluster maps were exported as NIfTI files into Mango brain visualisation software (http://ric.uthscsa.edu/mango/), and were overlain onto a canonical anatomical T1 brain template (Colin27_T1_seg_MNI.nii available from http://www.brainmap.org/ale/). A single-subject template was used to promote clarity given the multiple image presentation of the ALE results. Logical overlays were used within the Mango software to carry out overlap analysis.

3. Results

3.1. All studies

The cluster results of the ALE meta-analysis for all prediction error studies are listed in Table 2, and presented visually in the panel of brain-slices in Fig. 1. For all tables presenting cluster results, the size of each cluster is given in mm^3 together with the coordinates and level of the maximum ALE value that indicates the relative effect size for each extrema within each meta-analysis. In large clusters providing several peaks of activation, the foci of peak

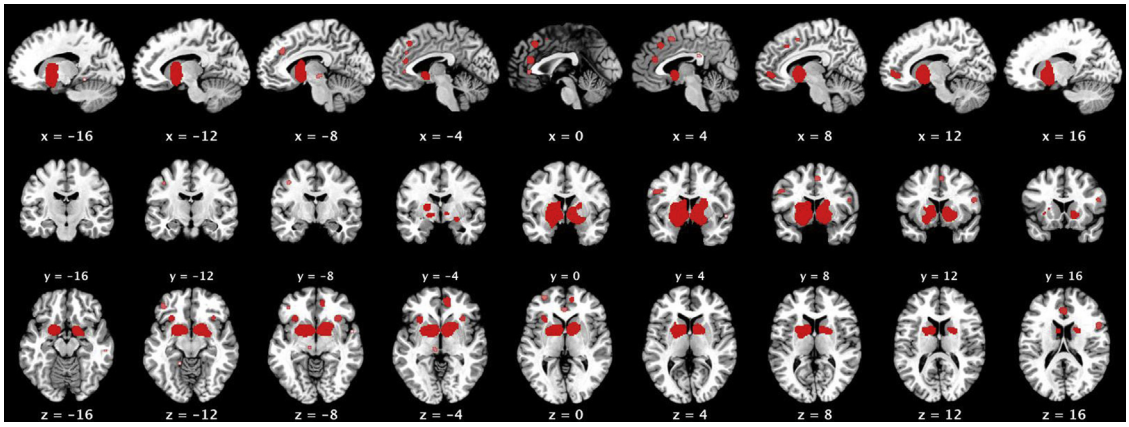

Fig. 1. Results of the ALE meta-analysis for all prediction error studies. All brain-panel imaging figures show representative slices in sagittal (top), coronal (middle) and axial (bottom) views with MNI planar coordinates given below each image.

activations are listed separately (e.g. left hand column in Table 2) and the cluster size (right hand column) is left as blank.

In the analysis of all reinforcement learning studies (Table 2 and Fig. 1) the most significant features are the large bilateral basal ganglia clusters with peaks in dorsal and ventral putamen, dorsal caudate and pallidum with the clusters also activating areas of thalamus, amygdala and hypothalamus. Striatal activation appears equally strong in the DS and VS, in left and right hemispheres and encompasses bilateral activation of the NAcc. Separate small basal ganglia clusters are centred on right and left claustrum (this latter cluster also extends to include a peak in the left insula). Beyond the basal ganglia, there are 24 further areas of activation across the brain, with larger clusters observed in the right cingulate cortex, and bilateral frontal cortex including the medial, inferior and superior frontal gyri. There is no clear effect of laterality across all brain regions; most non-striatal clusters are found in only one hemisphere but the distribution of these is balanced between left and right.

3.2. Instrumental and Pavlovian studies

Table 3 and Fig. 2 show the results of the ALE prediction error analyses for instrumental and Pavlovian studies separately. Distinct activation patterns for the two forms of reinforcement learning can be seen. The ALE analysis for instrumental studies show a single large bilateral cluster in the basal ganglia covering dorsal and ventral putamen, caudate, pallidum and NAcc. Further clusters are seen in the right and left claustrum.

The Pavlovian analysis indicates large bilateral basal ganglia clusters with a significant lateral effect of stronger activation in the left striatum. In the left hemisphere, activation encompasses peaks in the ventral and dorsal putamen, dorsal caudate and (extending dorsally to $z = 14$) caudate body, and ventrally (to $z = -28$) to include the parahippocampal gyrus. In the right hemisphere the striatal cluster only extends dorsally from $z = -5$ (caudate head) to $z = 14$ (putamen), with the cluster activation peaking in dorsal caudate. Notably, in both hemispheres the Pavlovian PE cluster in ventral striatum lies adjacent to the NAcc with only marginal overlap (1–2 mm at certain points) with the 1200 mm^3 cuboid space enclosing the NAcc as defined by the coordinates given in Ahsan et al. (2007). Given the wide margin of error in the use of the cuboid space to delineate the NAcc, it is considered unlikely that this marginal overlap indicates any degree of significant NAcc activation in the Pavlovian studies.

The overlap analysis shown in Fig. 2 indicates that instrumental basal ganglia prediction error activation is more widespread than that observed for the Pavlovian studies, which lie to the lateral extremes of the instrumental clusters. Beyond the basal ganglia, there are a further 11 Pavlovian and 13 instrumental clusters, found in both cases predominantly in the frontal cortex around the frontal gyrus, and in the cingulate cortex. However, with the exception of a small degree of overlap in the right cingulate gyrus, all PE clusters outside of the basal ganglia are specific to either instrumental or Pavlovian conditions.

Table 4 and Fig. 3 show the results of the ALE subtraction analysis for instrumental and Pavlovian prediction error studies. Whereas instrumental-Pavlovian PE activation lies solely within the bilateral basal ganglia (primarily ventral striatum), Pavlovian-instrumental clusters lie entirely beyond the basal ganglia, in the right cingulate and medial frontal gyri, left middle frontal gyrus and left insula. Thus the subtraction analysis confirms the earlier finding of stronger ventral striatal PE activation for instrumental compared to Pavlovian studies. It should be noted that the peak maxima shown in Table 4 may not correspond to any of those shown in Table 3 as the subtraction analyses may give rise to new maxima.

To see whether the counter intuitive finding of activation in the NAcc for instrumental, but not Pavlovian, was reliable, we conducted a GingerALE meta-analysis of 18 striatal ROI studies, excluded from the whole brain meta-analysis. GingerALE can be used in this way so long as the ROIs cover a similar volume and are not mixed with whole brain studies (e.g. Chen et al., 2011). 10 of these ROI studies examined instrumental learning (29 foci) and nine Pavlovian (39 foci; n.b. one study looked at both learning types). 5 studies looked at PE for punishment only, 5 for combined reward and punishment and the remaining 10 at RPE. 17 studies used an ROI that included the ventral striatum, of which 7 of these used a ROI that extended to include dorsal striatum. The results from this supplementary meta-analysis also showed that the clusters for the Pavlovian studies were to be found in both hemispheres but lay adjacent to, and did not encompass the NAcc. For the instrumental studies a considerably larger VS cluster was observed which included all of the NAcc in the right but not left hemisphere. This result indicates that the result arising from the whole brain meta-analysis is reliable.

3.3. Reward and punishment studies

Table 5 and Fig. 4 show the results of the ALE prediction error analyses for studies involving rewarding and punishing reinforcers. Despite the greater number of input foci for the reward condition (262) compared to the punishment condition (71), the activation pattern for reward studies is more concentrated as shown by the smaller number of cluster extrema in Table 5. While some discrepancy in the cluster activation patterns might be expected between

Table 3
Detailed ALE cluster results for instrumental and Pavlovian prediction error studies.

	L/R	*x*	*y*	*z*	ALE (10^{-3})	Size	
Instrumental							
Claustrum	L	−10	6	−6	49	16,528	Cluster encompasses NAcc
Caudate (VS)	R	12	8	−4	47		
Caudate (DS)	L	−10	6	8	26		
Putamen (DS)	R	22	4	12	20		
Putamen (DS)	L	−24	6	6	18		
Putamen (DS)	L	−22	16	4	12		
Anterior cingulate		0	34	18	16	552	
Medial frontal gyrus		0	28	44	20	520	
Cingulate gyrus	L	−6	34	30	16	336	
Middle frontal gyrus	L	−32	54	0	16	328	
Claustrum	L	−32	22	−8	14	320	
Claustrum	R	34	22	−8	14	296	
Inferior parietal lobule	R	46	−40	54	15	264	
Medial frontal gyrus	R	26	38	28	14	208	
Superior frontal gyrus	L	−24	32	48	14	208	
Middle temporal gyrus	R	58	−26	−18	15	200	
Superior temporal gyrus	R	56	4	−8	14	176	
Precuneus	R	36	−72	48	16	112	
Anterior cingulate	R	8	50	−4	12	72	
Superior occipital gyrus	R	38	−82	34	18	72	
Superior frontal gyrus	L	−18	0	68	12	72	
Pavlovian							
Putamen (DS)	L	−18	2	8	19	3120	Cluster borders but does not encompass NAcc
Putamen (VS)	L	−20	8	−14	16		
Caudate (VS)	L	−10	8	2	14		
Parahippocampal gyrus	L	−20	2	−24	12		
Caudate (VS)	R	10	8	2	21	1592	Cluster extends into dorsal striatum; borders but does not encompass NAcc
Superior frontal gyrus	R	4	10	50	17	888	
Medial frontal gyrus	L	−6	4	54	10		
Cingulate gyrus	R	6	24	38	13	432	
Insula	L	−36	28	6	12	384	
Claustrum	L	−30	22	−2	10		
Middle frontal gyrus	L	−28	16	58	11	296	
Inferior frontal gyrus	R	54	14	18	14	264	
Pallidum	R	26	−10	−8	10	232	
Putamen (VS)	R	28	−2	−12	10		
Red nucleus	R	6	−20	−10	13	232	
Anterior cingulate	R	12	42	−4	12	224	
Cerebellum	R	34	−44	−24	11	184	
Middle frontal gyrus	R	30	24	54	10	104	
Superior frontal gyrus	R	24	22	56	9		
Thalamus	L	−8	−28	−4	10	72	

Abbreviations: DS, dorsal striatum; VS, ventral striatum; NAcc, nucleus accumbens; L, left hemisphere; R, right hemisphere.

reward and punishment studies due to the larger number of studies (more foci) in the reward group, this is insufficient to explain the exceptionally low relative level of striatal activation for punishment studies. While two large bilateral basal ganglia RPE clusters encompass both dorsal and ventral caudate and putamen (including NAcc), the analysis of punishment studies shows only two small left hemisphere striatal clusters (peaking in left dorsal and ventral putamen). Notably, no PE cluster activation is found for punishment

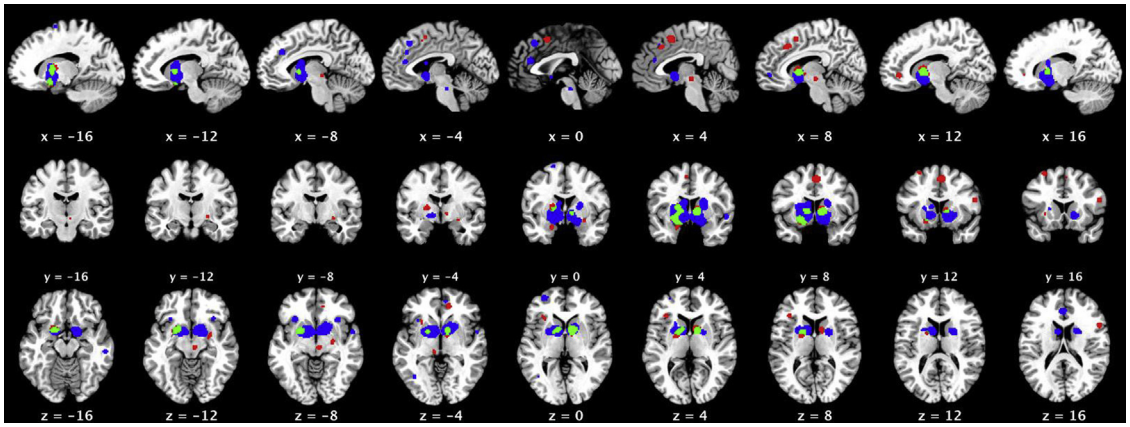

Fig. 2. Results of the ALE meta-analysis for instrumental (blue) and Pavlovian (red) prediction error studies. The overlap of the two analyses is shown in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4
Detailed ALE cluster results for instrumental and Pavlovian prediction error studies: subtraction analyses.

	L/R	*x*	*y*	*z*	ALE (10^{-3})	Size	
Instrumental – Pavlovian							
Caudate (VS)	R	8	10	−14	2948	1872	Cluster encompasses NAcc
Caudate (VS)	R	10	14	−10	2820		
Pallidum	L	−8	0	−10	2748	1472	Cluster extends into ventral striatum including NAcc
Caudate (DS)	L	−8	4	18	1943	88	
Pavlovian – Instrumental							
Cingulate gyrus	R	2	12	47	3156	840	
Medial frontal gyrus	R	0	8	52	2989		
Middle frontal gyrus	L	−28	20	56	1999	296	
Middle frontal gyrus	L	−30	16	56	1960		
Insula	L	−34	27	5	1957	120	

Abbreviations: DS, dorsal striatum; VS, ventral striatum; NAcc, nucleus accumbens; L, left hemisphere; R, right hemisphere.

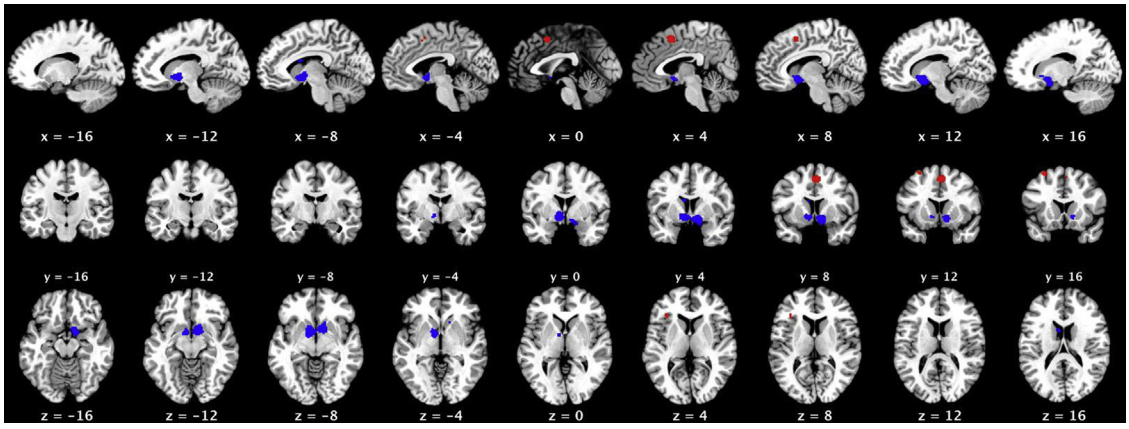

Fig. 3. Results of the ALE subtraction analysis for [instrumental-Pavlovian] (blue) and [Pavlovian-instrumental] (red) prediction error studies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

studies in the right striatum. In contrast, no clear effect of laterality is observed for striatal reward clusters.

Beyond the basal ganglia, PE clusters for both reward and punishment studies are widely distributed with no clear effect of laterality. For both groups the largest clusters are seen in the frontal and cingulate cortices, with activation seen at a lower level in the temporal and occipital gyri for RPE and fusiform and temporal gyri, and superior parietal lobule for APE. The overlap analysis reveals shared reward and punishment PE activation only within the striatal clusters and in the medial anterior cingulate. Aside from these areas cluster activations are specific to each of the conditions.

The ALE subtraction analyses of reward and punishment studies shown in Table 6 and Fig. 5 confirms the pattern of activation described above. There is greater PE activity for rewarding tasks compared to those involving punishment in the basal ganglia, with a large reward cluster in each hemisphere that encompasses the NAcc. The right side cluster covers both DS and VS whereas the left side cluster is predominantly ventral, with $z < 0$. Notably, this large cluster has extrema in the right hypothalamus and pallidum despite these regions not appearing as peaks in the reward cluster table. A further small left hemisphere dorsal caudate cluster peaks at $z = 16$ and 18. In contrast the punishment–reward analysis reveals no basal ganglia clusters, but three small clusters are found in regions, which show no activation for reward studies, namely left thalamus and insula, and the middle frontal gyrus.

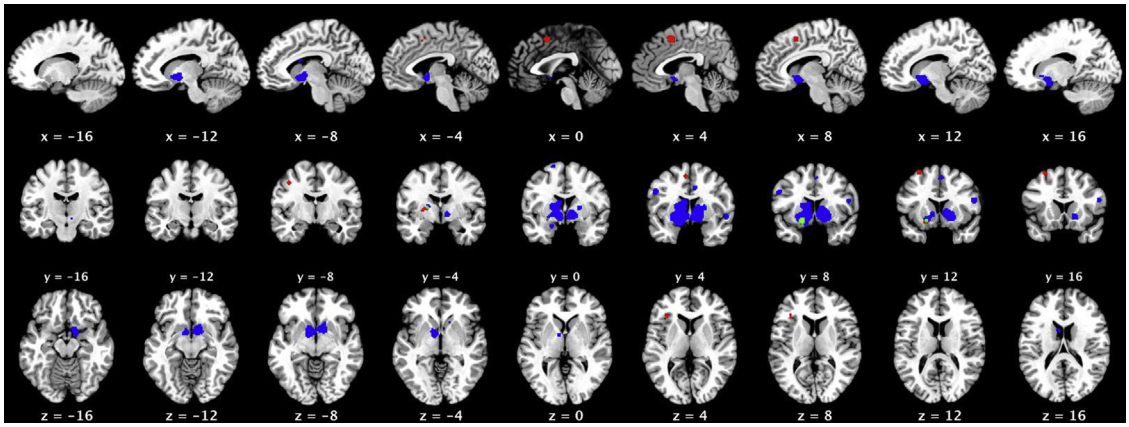

Fig. 4. Results of the ALE meta-analysis for reward (blue) and aversive prediction error studies (red). The overlap of the two analyses is shown in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 5
Detailed ALE cluster results for reward and aversive prediction error studies.

	L/R	*x*	*y*	*z*	ALE (10^{-3})	Size	
Reward							
Claustrum	L	−10	6	−6	35	7440	Cluster encompasses NAcc
Putamen (VS)	L	−18	8	−12	24		
Putamen (DS)	L	−18	2	8	22		
Caudate (DS)	L	−10	4	10	20		
Parahippocampal gyrus	L	−20	2	−22	12		
Caudate (VS)	R	10	6	−2	41	5360	Cluster encompasses NAcc
Putamen (DS)	R	22	2	6	14		
Inferior frontal gyrus	R	54	12	18	17	424	
Cingulate gyrus	R	6	26	38	14	384	
Precentral gyrus	L	−52	6	30	15	328	
Superior frontal gyrus	L	−24	32	48	14	248	
Anterior cingulate	R	8	50	−4	12	208	
Middle temporal gyrus	R	58	−26	−18	15	200	
Red nucleus	R	6	−20	−10	13	176	
Superior temporal gyrus	R	56	4	−8	14	176	
Cingulate gyrus	R	8	4	36	14	176	
Substantia nigra	L	−10	−20	−8	13	168	
Superior frontal gyrus	R	4	12	52	13	152	
Superior occipital gyrus	R	38	−82	34	18	104	
Cingulate gyrus	R	2	32	20	13	96	
Superior frontal gyrus	L	−18	0	68	12	88	
Punishment							
Putamen (VS)	L	−18	8	−14	13	400	Ventral striatum cluster bordering but not including NAcc
Anterior cingulate	R	2	34	18	12	384	
Thalamus	L	−6	−26	−2	11	376	
Cerebellum	R	34	−44	−24	11	320	
Insula	L	−36	28	6	11	320	
Middle frontal gyrus	L	−30	12	60	10	224	
Middle frontal gyrus	R	30	24	54	10	208	
Superior frontal gyrus	R	24	22	56	9		
Pallidum	L	−20	−4	6	9	160	Predominantly dorsal striatum cluster
Putamen	L	−28	−6	2	9		
Medial frontal gyrus	R	8	42	26	9	128	
Medial frontal gyrus	R	12	40	34	9		
Fusiform gyrus	R	46	−76	−10	10	72	
Inferior frontal gyrus	R	36	30	2	10	72	
Medial frontal gyrus	R	12	58	6	10	64	
Inferior temporal gyrus	L	−52	−26	−18	9	56	
Anterior cingulate	L	−6	42	−12	9	56	
Middle frontal gyrus	L	−40	40	−10	9	56	
Superior temporal gyrus	L	−50	−22	−4	9	56	
Lingual gyrus	L	−20	−96	0	9	56	
Lingual gyrus	R	18	−72	0	9	56	
Superior frontal gyrus	L	−12	56	30	9	56	
Angular gyrus	R	50	−68	32	9	56	
Superior parietal lobule	L	−24	−56	44	9	56	
Precentral gyrus	L	−40	−8	44	9	56	
Medial frontal gyrus	L	−6	4	54	9	56	

Abbreviations: DS, dorsal striatum; VS, ventral striatum; NAcc, nucleus accumbens; L, left hemisphere; R, right hemisphere.

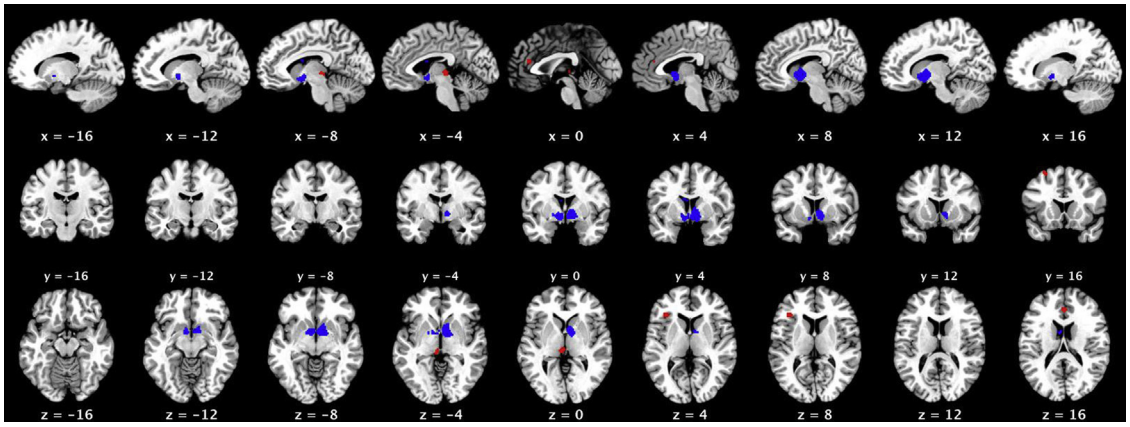

Fig. 5. Results of the ALE subtraction analysis for Reward-Punishment (blue) and Punishment-Reward (red) prediction error studies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 6
Detailed ALE cluster results for reward and aversive prediction error studies: subtraction analyses.

	R	*x*	*y*	*z*	ALE (10^{-3})	Size	
Reward – punishment							
Hypothalamus	R	10	−3	−7	1957	2424	Cluster encompasses dorsal striatum and ventral striatum including NAcc
Anterior cingulate	R	4	5	−11	1957		
Pallidum	R	10	2	−6	1957		
Thalamus	R	10	−4	−2	1957		
Pallidum	R	12	0	−2	1957		
Pallidum	R	12	2	−12	1957		
Anterior cingulate	L	−2	2	−9	1957	776	Cluster encompasses ventral striatum including NAcc, bordering dorsal striatum at $z < 1$
Pallidum	L	−7	−2	−8	1957		
Pallidum	L	−18	0	−6	1957		
Caudate	L	−5	1	16	1957	88	Dorsal striatum cluster
Caudate	L	−10	3	18	1957		
Punishment – Reward							
Thalamus	L	−3	−25	1	2297	296	
Thalamus	L	−8	−29	−2	1861		
Insula	L	−38	26	8	2050	296	
Insula	L	−34	28	8	1977		
Anterior cingulate	R	1	36	15	1765	128	
Middle frontal gyrus	L	−28	16	56	1789	80	
Middle frontal gyrus	L	−30	15	60	1719		

Abbreviations: DS, dorsal striatum; VS, ventral striatum; NAcc, nucleus accumbens; L, left hemisphere; R, right hemisphere.

3.4. Summary of results

Taking all studies together (i.e. combining all types of instrumental and Pavlovian PEs including both rewards and punishments) we found a large cluster of prediction error related activation in the striatum, which embraced both ventral and dorsal regions. We also found activation that extends to medial frontal structures including pregenual and antero-medial cingulate cortex (see Fig. 4), which have anatomical connections to the midbrain and/or striatum. The results showed that prediction error was coded in diverse regions throughout the cerebral cortex, (we noted 24 areas; see, Table 2), predominantly in anterior rather than posterior regions. This finding concurs with the review of the animal literature by Schultz and Dickinson (2000).

When comparing areas of activation for instrumental and Pavlovian conditioning, we noted considerable overlap in dorsomedial striatum bilaterally and in left ventral putamen in the VS complex. Surprisingly there was no overlap of prediction error coding for both forms of learning in NAcc in either the left or right hemispheres. NAcc coded prediction error only in instrumental learning but not Pavlovian. Outside of the basal ganglia there was also little overlap of PE coding for Pavlovian and instrumental learning. This finding was also strongly supported by the subtraction analysis.

The meta-analysis also indicated that reward and aversive prediction errors are coded throughout the brain by broadly segregated neural networks with minimal integration between RPE and APE's in the ventral striatum and antero-medial cingulate cortex. This suggests possible differences in the processing of aversive and appetitive reinforcers (Fujiwara et al., 2009; Grabenhorst and Rolls, 2011).

4. Discussion

4.1. RPE in instrumental vs. Pavlovian reinforcement learning

According to the actor–critic model of reinforcement learning, the critic calculates the discrepancy between the expected value of a stimulus and its actual value in both Pavlovian as well as instrumental learning (McClure et al., 2004; van der Meer and Redish, 2011). If we assume a single critic then the ALE maps for the critic should be similar for both forms of reinforcement learning. In the striatum, this similarity is most apparent in the dorsomedial region rather than ventral striatum, a finding that concurs with the early actor–critic models of reinforcement learning (reviewed in Joel et al., 2002). More recent reviews have argued that prediction error and hence the critic, are computed in ventral striatum (McClure et al., 2004; van der Meer and Redish, 2011; Doll et al., 2012). However, in the current study, PE in ventral striatum during Pavlovian conditioning was confined to left putamen only, with no activity in the NAcc. Conversely, during instrumental learning, PE did correlate significantly with activation throughout the NAcc, ventral putamen and ventral caudate on both left and right sides, a finding which was supported by the subtraction analysis (instrumental-Pavlovian).

These results suggest that the ventral striatum is not operating as a critic for use in both instrumental and Pavlovian reinforcement learning in humans. However, it does not rule out the possibility that the functional organisation of the striatum follows the actor–critic model, but without a direct one-to-one mapping. It might be that there is more than one actor–critic implemented in the striatum (Doya et al., 2002; Baldassarre, 2002) or that there are different actor–critic systems, such as one for model-based and another for model-free processes of instrumental behaviour (see van der Meer and Redish, 2011; Bornstein and Daw, 2011; Doll et al., 2012 for a discussion). Alternatively, different regions in the striatum may each contribute to a 'distributed' critic function, perhaps with the NAcc contributing more to the estimation of prediction error in the case of instrumental learning (Redgrave and Gurney, 2006).

Two adaptations of the conventional view could also account for these findings. McClure et al. (2003) present a computational model in which PE is used to assign incentive salience by updating not only predictions of future rewarding events (i.e. the conventional critic) but also biasing action choice (the conventional actor). A second adaptation is offered by van der Meer and Redish (2011) who suggest that the critic also processes anticipated state values, more specifically those expected by the agent to occur as a result of their choice of action. Both explanations assume greater activation across the ventral striatum when actions are required (i.e. in instrumental learning) as opposed to when the animal is passive (Pavlovian).

It might be argued that the concept of ventral striatum operating simply as a 'critic' is fundamentally flawed since there is a conflux of information from distinct sources converging onto this region (Humphries and Prescott, 2010). Others have provided evidence to show that this area plays a key role in action selection and Pavlovian-instrumental transfer (Nicola, 2007; Yin and Knowlton, 2006). Berridge (2007, 2012) proposes that the ventral striatum is involved more with incentive salience, or the motivational aspect of 'wanting' (i.e. adjusting the value of CS and US to fit current motivational state), rather than of 'cognitive wanting' which refers to the calculation of prediction error. Ventral striatal activity would thus be expected to vary according to the motivational value of both the CS and US. This could account for the observed absence of NAcc PE involvement in Pavlovian reinforcement learning, as opposed to that observed for instrumental learning where the agent must engage with the environment to make a response (e.g. the response will occur only if the agent is motivated to act).

4.2. RPE coding for reward and punishment

In Section 3.4, we note the largely separate systems involved in calculating RPE and APE. We identified activations in the midbrain for reward prediction errors, consistent with those areas reported in animal studies that showed increased spiking of DA neurons for the unexpected occurrence of rewards (Hollerman and Schultz, 1998; Bayer and Glimcher, 2005). Furthermore, widespread activation in the ventral striatum was observed for reward prediction errors while aversive prediction errors were encoded in a restricted cluster in the left ventral striatum. This finding concurs with neuroanatomical evidence that shows dense connections between the major axonal pathways of dopaminergic neurons, which putatively code for RPE, and the striatum (Arbuthnott and Wickens, 2007).

Robust APE activation clusters were detected in both the habenula and insula cortex, with the habenula result being consistent with the findings reported by Salas et al. (2010). This region has been implicated in several emotional and cognitive functions including pain, learning and attention (Hikosaka, 2010). In contrast, insula cortex is broadly acknowledged as viscerosensory cortex, and is implicated in the mapping of internal bodily states (including pain and taste) and in the representation of emotional arousal and feelings (Singer et al., 2009).

Both reward and aversive prediction error clusters were found in anterior cingulate cortex (ACC) in keeping with the results of electrophysiology studies in primates (Amiez et al., 2006; Matsumoto and Hikosaka, 2007), and EEG studies in humans (Holroyd and Coles, 2002; Nieuwenhuis et al., 2005; Oliveira et al., 2007). There were, however, notable regional differences within the ACC. Pregenual ACC was active only for reward prediction errors whereas antero-medial ACC activity correlated with both reward and aversive prediction errors. Antero-medial ACC has previously been reported to play a role in negative affectivity and pain (Price, 2000; Farrell et al., 2005; Lancaster et al., 2000; Peyron et al., 2000; Vogt, 2005), and signalling aversive rather than reward prediction errors (Shackman et al., 2011). It has also been associated with conflict-monitoring (Botvinick et al., 2004) or error detection (Carter et al., 1998), both of which might be alternative explanations to our finding that this region was coding RPE as well as APE.

It is also acknowledged that observation of distinct systems involved in RPE and APE might also result from a confound with the different action requirements of the learning tasks, namely 'go', or approach behaviours in rewarded trials as opposed to 'no-go' or avoid behaviours in punished trials (Guitart-Masip et al., 2011).

4.3. fMRI modelling and the dopamine reward prediction error hypothesis

Nearly all of the 35 studies in this review make reference to the role of dopamine (DA) in computing PE (the dopamine reward prediction hypothesis) as described in the seminal papers by Read Montague (e.g. Montague et al., 1996) and Wolfram Schultz (e.g. Schultz et al., 1997). It was typically inferred that these fMRI-based studies provide a window onto the functionality of the DA system. Indeed, six of the studies reviewed used the putative fMRI measure of PE to compare the results of healthy controls to patients with dysfunctional DA systems (Kumar et al., 2008; Murray et al., 2008; O'Sullivan et al., 2011; Schonberg et al., 2010) or to explore the effects of medications used with these patients which act on their DA system (Jocham et al., 2011; Pessiglione et al., 2006).

Our meta-analysis synthesised foci from different studies derived from analyses in which the estimate of PE had been regressed onto BOLD activity time-locked to US presentation. Such an analysis does not exclusively rule out other explanations for the observed BOLD signal activations such as surprise, shift of selective attention or the influence of reward magnitude that correlates with PE. By way of example, Roesch et al. (2010) used several criteria to investigate whether neuronal activity, that correlated with unexpected reward in the amygdala, was actually RPE, outcome expectancy or, as they concluded, shift of attention.

However, the studies undertaken by John O'Doherty and his collaborators (e.g. O'Doherty et al., 2003; Tobler et al., 2006; Valentin et al., 2009) made explicit that the 'full range of TD error-related PE responses' was required to confirm the DA prediction error hypothesis. The criteria they offered included an elevated response at the time of reward presentation in early learning (criterion 1), transfer of the PE response to the timing of CS presentation once the CS-US association is established (criterion 2), at which stage a large positive PE should result from any unexpected reward (criterion 3) and a large negative PE when an expected reward is omitted (criterion 4).

Only criterion 1 is directly covered by this meta-analysis but many of the studies have also explored criteria 2–4 typically using a ROI approach. Several studies (O'Doherty et al., 2003; Tobler et al., 2006; Brovelli et al., 2008; Morris et al., 2011; Pessiglione et al., 2006; McClure et al., 2003; Gläscher et al., 2009; Ramnani et al., 2004) reported deactivation of the BOLD signal relative to a baseline in striatum, predominantly ventral striatum, following unexpected reward omission (criterion 4) and increased activation for an unexpected reward (criterion 3).

With regard to criterion 2, both O'Doherty et al. (2003) and Tobler et al. (2006) found that PE striatal activation transferred from the timing of the US in early trials to that of the CS in later trials, although this was not found in the study by Valentin et al. (2009). Further evidence also suggests that criterion 2 may not be observed in the striatum. Although Li et al. (2011a) reported activity in striatum that met criterion 2 this was only true for trial-and-error learning (i.e. the typical experimental paradigm used in instrumental learning) but not when symbolic knowledge (e.g. information) is included in the learning task. Similarly, Seger et al. (2010) in explaining the lack of difference between negative and positive feedback in late but not early trials comment 'the effect of feedback valence may depend on factors such as . . .. the utility of the feedback . . ..with positive feedback being more useful early in learning, and negative feedback more useful as learning progresses and errors are rarer'.

All of the studies included in this meta-analysis have assumed the BOLD signal used in fMRI to be an indirect measure of neuronal

activity. However, DA receptors are found on some micro-vessels in the brain and DA release can itself have a direct effect on blood perfusion (Choi et al., 2006). Thus, if we assume that the BOLD signal in these studies reflects the release of DA, then the observed results might reflect not only neuronal activity but also the direct effects of DA on vascular mechanisms. Thus it is probably expedient at this stage to be cautious in concluding that the patterns of active foci emerging from regression analyses of PE onto an fMRI trace, time-locked to the US, offer a robust test of the dopamine reward prediction error hypothesis in humans. None-the-less, the results of this meta-analysis enable us to question the accepted mindset regarding the mapping of the actor–critic elements of computational learning models onto dorsal and ventral striatum respectively.

4.4. Caveats

This meta-analysis has several methodological caveats. Many fMRI studies of prediction error in humans have used striatal ROI analysis given the findings of electrophysiological studies in animals. In order to avoid bias in our results, all ROI studies were excluded from the main meta-analysis, although we have made reference to some in order to corroborate the distinct VS findings for instrumental and Pavlovian learning. It was also impossible to separate studies of instrumental learning into those requiring goal based responding ('model-based' algorithms) and those requiring automatic or habit type responding ('model-free' algorithms). There is considerable theoretical and empirical support for the involvement of distinct brain areas (in particular frontal and striatal regions) in model-based and model-free decision making (Humphries and Prescott, 2010; van der Meer and Redish, 2010). Many of the studies in this meta-analysis comprised simple discriminative learning paradigms where early trials may require a model-based approach, with later trials a model-free approach.

As with other ALE meta-analyses, we ignored the sign of the fMRI BOLD signal. Some studies reported negative activations for aversive prediction errors, but these were few in number and the effect size was small. Furthermore, given the limited number of studies reporting reward magnitude, uncertainty or probability of loss relative to gain, these measures were not included in our analysis although they do appear to influence the magnitude of the BOLD signal. A more general limitation of Ginger ALE is its current inability to take into account effect sizes (e.g. between studies or different brain regions), or the extent of activation for each input cluster. This may skew the results as some studies may report a single peak cluster activation whereas others may report several peak activations within the same cluster. Both may have the same magnitude and extent of activation but the study reporting several coordinates will have more power in the ALE analysis.

Conflict of interest

All authors report no conflict of interest.

Acknowledgements

The authors would particularly like to acknowledge the valuable help and advice provided at several stages throughout this study by Dr. Angie Laird at the Research Imaging Centre, University of Texas, San Antonio. We would also like to thank the following colleagues for their critical comments of earlier drafts: Profs. Peter Dayan, Kevin Gurney, Kent Berridge and the following for kindly providing data not previously in the public domain: Noreen O'Sullivan, Philippe Tobler, Mathias Pessiglione, Thorsten Kahnt, Gerhard Jocham.

Appendix A.

Studies included in the meta-analysis.

Brovelli, A., Laksiri, N., Nazarian, B., Meunier, M., Boussaoud, D., 2008. Understanding the neural computations of arbitrary visuomotor learning through fMRI and associative learning theory. Cereb. Cortex 18(7), 1485–1495.
Cohen, M.X., 2007. Individual differences and the neural representations of reward expectation and reward prediction error. Soc. Cogn. Affect. Neurosci. 2(1), 20–30.
Delgado, M.R., Li, J., Schiller, D., Phelps, E.A., 2008. The role of the striatum in aversive learning and aversive prediction errors. Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 363(1511), 3787–800.
Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66(4), 585–595.
Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex 19(2), 483–495.
Gradin, V.B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., et al., 2011. Expected value and prediction error abnormalities in depression and schizophrenia. Brain 134(6), 1751–1764.
Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J. Neurosci. 26(32), 8360–8367.
Howard-Jones, P.A., Bogacz, R., Yoo, J.H., Leonards, U., Demetriou, S., 2010. The neural mechanisms of learning from competitors. NeuroImage 53(2), 790–799.
Jocham, G., Klein, T.A., Ullsperger, M., 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J. Neurosci. 31(5), 1606–1613.
Kahnt, T., Grueschow, M., Speck, O., Haynes, J.-D., 2011. Perceptual learning and decision-making in human medial frontal cortex. Neuron 70(3), 549–559.
Kahnt, T., Park, S.Q., Cohen, M.X., Beck, A., Heinz, A., Wrase, J., 2009. Dorsal striatal-midbrain connectivity in humans predicts how reinforcements are used to guide decisions. J. Cogn. Neurosci. 21(7), 1332–1345.
Kim, H., Shimojo, S., O'Doherty, J.P., 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. 4(8), e233.
Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., Steele, J.D., 2008. Abnormal temporal difference reward-learning signals in major depression. Brain 131(8), 2084–2093.
Landmann, C., Dehaene, S., Pappata, S., Jobert, A., Bottlaender, M., Roumenov, D., Le Bihan, D., 2007. Dynamics of prefrontal and cingulate activity during a reward-based logical deduction task. Cereb. Cortex 17(4), 749–759.
Li, J., Delgado, M.R., Phelps, E.A., 2011a. How instructed knowledge modulates the neural systems of reward learning. Proc. Natl. Acad. Sci. U. S. A. 108(1), 55–60.
Li, J., McClure, S.M., King-Casas, B., Montague, P.R., 2006. Policy adjustment in a dynamic economic game. PLoS ONE 1(1), e103.
Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A., Daw, N.D., 2011b. Differential roles of human striatum and amygdala in associative learning. Nat. Neurosci. 14(10), 1250–1252.
McClure, S.M., Berns, G.S., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron 38(2), 339–246.
Morris, R.W., Vercammen, A, Lenroot, R., Moore, L., Langton, J.M., Short, B., Kulkarni, J., et al., 2011. Disambiguating ventral striatum

fMRI-related bold signal during reward prediction in schizophrenia. Mol. Psychiatry, 17, 280–289.
Murray, G.K., Corlett, P.R., Clark, L., Pessiglione, M., Blackwell, A.D., Honey, G., 2008. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol. Psychiatry 13, 267–276.
O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38(2), 329–337.
O'Sullivan, N., Szczepanowski, R., El-Deredy, W., Mason, L., Bentall, R.P., 2011. fMRI evidence of a relationship between hypomania and both increased goal-sensitivity and positive outcome-expectancy bias. Neuropsychologia 49(10), 2825–2835.
Park, S.Q., Kahnt, T., Beck, A., Cohen, M.X., Dolan, R.J., Wrase, J., Heinz, A., 2010. Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. J. Neurosci. 30(22), 7749–7753.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045.
Ramnani, N., Elliott, R., Athwal, B.S., Passingham, R.E., 2004. Prediction error for free monetary reward in the human prefrontal cortex. Cortex 23, 777–786.
Rodriguez, P.F., Aron, A.R., Poldrack, R.A., 2006. Ventral–striatal/nucleus–accumbens sensitivity to prediction errors during classification learning. Hum. Brain Mapp. 27(4), 306–313.
Schonberg, T., O'Doherty, J.P., Joel, D., Inzelberg, R., Segev, Y., Daw, N.D., 2010. Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. NeuroImage 49(1), 772–781.
Seger, C.A., Peterson, E.J., Cincotta, C.M., Lopez-Paniagua, D., Anderson, C.W., 2010. Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. NeuroImage 50(2), 644–656.
Seymour, B., Daw, N., Dayan, P., Singer, T., Dolan, R., 2007. Differential encoding of losses and gains in the human striatum. J. Neurosci. 27(18), 4826–4831.
Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., Dolan, R., 2005. Opponent appetitive–aversive neural processes underlie predictive learning of pain relief. Nat. Neurosci. 8(9), 1234–1240.
Spoormaker, V I, Andrade, K.C., Schröter, M.S., Sturm, A, Goya-Maldonado, R., Sämann, P.G., Czisch, M., 2011a. The neural correlates of negative prediction error signaling in human fear conditioning. NeuroImage 54(3), 2250–2256.
Spoormaker, Victor I, Schröter, M.S., Andrade, K.C., Dresler, M., Kiem, S.A., Goya-Maldonado, R., Wetter, T.C., et al., 2011b. Effects of rapid eye movement sleep deprivation on fear extinction recall and prediction error signaling. Hum. Brain Mapp. 33(10), 2362–2376.
Tanaka, S.C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., Doya, K., 2006. Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. Neural Netw. 19(8), 1233–1241.
Tobler, P.N., O'Doherty, J.P., Dolan, R.J., Schultz, W., 2006. Human neural learning depends on reward prediction errors in the blocking paradigm. J. Neurophysiol. 95(1), 301–310.
Valentin, V.V., O'Doherty, J.P., 2009. Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. J. Neurophysiol. 102(6), 3384–3391.

References

Ahsan, R.L., Allom, R., Gousias, I.S., Habib, H., Turkheimer, F.E., Free, S., Lemieux, L., et al., 2007. Volumes, spatial extents and a probabilistic atlas of the human basal ganglia and thalamus. Neuroimage 38 (2), 261–270.
Amiez, C., Joseph, J.P., Procyk, E., 2006. Reward encoding in the monkey anterior cingulate cortex. Cereb. Cortex 16 (7), 1040–1055.
Arbuthnott, G.W., Wickens, J., 2007. Space, time and dopamine. Trends Neurosci. 30 (2), 62–69.
Baldassarre, G., 2002. A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviors. J. Cogn. Syst. Res. 3 (1), 5–13.
Balleine, B., Daw, N., O'Doherty, J.,2008. Multiple forms of value learning and the function of dopamine. In: Neuroeconomics. Academic Press, United Kingdom, pp. 367–387.
Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47 (1), 129–141.
Barto, A.G., 1995. Adaptive critics and the basal ganglia. In: Houk, J.C., Davis, J., Beiser, D. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, MA, pp. 215–232.
Berridge, K.C., 2007. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berlin) 191 (3), 391–431.
Berridge, K.C., 2012. From prediction error to incentive salience: mesolimbic computation of reward motivation. Eur. J. Neurosci. 35 (7), 1124–1143.
Bornstein, A.M., Daw, N.D., 2011. Multiplicity of control in the basal ganglia: computational roles of striatal sub-regions. Curr. Opin. Neurobiol. 21 (3), 374–380.
Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8 (12), 539–546.
Bush, R.R., Mosteller, F.A., 1951. Mathematical model for simple learning. Psychol. Rev. 58, 313–323.
Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280 (5364), 747–749.
Chen, C.-H., Suckling, J., Lennox, B.R., Ooi, C., Bullmore, E.T., 2011. A quantitative meta-analysis of fMRI studies in bipolar disorder. Bipolar Disord. 13 (1), 1–15.
Choi, J.K., Chen, Y.I., Hamel, E., Jenkins, B.G., 2006. Brain hemodynamic changes mediated by dopamine receptors: role of the cerebral microvasculature in dopamine-mediated neurovascular coupling. Neuroimage 30, 700–712.
Dayan, P., 2009. Goal-directed control and its antipodes. Neural Netw. 22, 213–219.
Dayan, P., Abbott, L.F., 2001. Theoretical Neuroscience: Computational and Mathematical modelling of Neural Systems. MIT Press, Cambridge, MA.
Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8, 1704–1711.
Doll, B.B., Simon, D.A., Daw, N.D., 2012. The ubiquity of model-based reinforcement learning. Curr. Opin. Neurobiol. (September).
Doya, K., Samejima, K., Katagiri, K., Kawato, M., 2002. Multiple model-based reinforcement learning. Neural Comput. 14 (6), 1347–1369.
Eickhoff, S.B., Bzdok, D., Laird, A.R., Roski, C., Caspers, S., Zilles, K., Fox, P.T., 2011. Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. Neuroimage 57 (3), 938–949.
Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp. 30 (9), 2907–2926.
Farrell, M.J., Laird, A.R., Egan, G.F., 2005. Brain activity associated with painfully hot stimuli applied to the upper limb: a meta-analysis. Hum. Brain Mapp. 25 (1), 129–139.
Fujiwara, J., Tobler, P.N., Taira, M., Iijima, T., Tsutsui, K.-I., 2009. Segregated and integrated coding of reward and punishment in the cingulate cortex. J. Neurophysiol. 101 (6), 3284–3293.
Glimcher, P.W., 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. U. S. A. 108 (Suppl.), 15647–15654.
Guitart-Masip, M., Fuentemilla, L., Bach, D.R., Huys, Q.J., Dayan, P., Dolan, R.J., Duzel, E., 2011. Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J. Neurosci. 31, 7867–7875.
Grabenhorst, F., Rolls, E.T., 2011. Value, pleasure and choice in the ventral prefrontal cortex. Trends Cogn. Sci. 15 (2), 56–67.
Haber, S.N., McFarland, N.R., 1999. The concept of the ventral striatum in nonhuman primates. Ann. N.Y. Acad. Sci. 877, 33–48.
Hikosaka, O., 2010. The habenula: from stress evasion to value-based decision-making. Nat Rev Neurosci 11 (7), 503–513.
Hollerman, J.R., Schultz, W., 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. 1 (4), 304–309.
Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109 (4), 679–709.
Houk, J.C., Adams, J.L., Barto, A.G., 1995. A model of how the basal ganglia generate and use reward signals that predict reinforcement. In: Houk, C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, pp. 249–270.
Humphries, M.D., Prescott, T.J., 2010. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog. Neurobiol. 90 (4), 385–417.
Joel, D., Niv, Y., Ruppin, E., 2002. Actor–critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. 15 (4–6), 535–547.
Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996. Reinforcement learning: a survey. J. Artif. Intell. Res. 4, 237–285.
Khamassi, M., 2005. Actor–critic models of reinforcement learning in the basal ganglia: from natural to artificial rats. Adapt. Behav. 13 (2), 131–148.

Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Turkeltaub, P.E., et al., 2005. ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. Hum. Brain Mapp. 25 (1), 155–164.
Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., et al., 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp. 10 (3), 120–131.
Liu, X., Hairston, J., Schrier, M., Fan, J., 2010. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. 35, 1219–1236.
Lohrenz, T., McCabe, K., Camerer, C.F., Montague, P.R., 2007. Neural signature of fictive learning signals in a sequential investment task. Proc. Natl. Acad. Sci. U. S. A. 104 (22), 9493–9498.
Mackintosh, N.J., 1975. A theory of attention: variations in the associability of stimuli with reinforcement. Psychol. Rev. 82, 276–298.
Martin, J.J., 2003. Neuroanatomy: Text and Atlas. McGraw-Hill, New York.
Matsumoto, M., Hikosaka, O., 2007. Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447 (7148), 1111–1115.
McClure, S.M., Daw, N.D., Montague, P.R., 2003. A computational substrate for incentive salience. Trends Neurosci. 26, 423–428.
McClure, S.M., York, M.K., Montague, P.R., 2004. The neural substrates of reward processing in humans: the modern role of FMRI. Neuroscientist 10 (3), 260–268.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning? J. Neurosci. 16 (5), 1936–1947.
Nicola, S.M., 2007. The nucleus accumbens as part of a basal ganglia action selection circuit. Psychopharmacology (Berlin) 191 (3), 521–550.
Nieuwenhuis, S., Slagter, H.A., von Geusau, N.J.A., Heslenfeld, D.J., Holroyd, C.B., 2005. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. Eur. J. Neurosci. 21 (11), 3161–3168.
Niv, Y., Schoenbaum, G., 2008. Dialogues on prediction errors. Trends Cogn. Sci. 12 (7), 265–272.
Oliveira, F.T.P., McDonald, J.J., Goodman, D., 2007. Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J. Cogn. Neurosci. 19 (12), 1994–2004.
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 (5669), 452–454.
O'Doherty, J.P., Hampton, A., Kim, H., 2007. Model-based fMRI and its application to reward learning and decision making. Ann. N. Y. Acad. Sci. 1104, 35–53.
Pearce, J.M., Hall, G., 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev. 87, 532–552.
Peyron, R., García-Larrea, L., Grégoire, M.C., Convers, P., Richard, A., Lavenne, F., Barral, F.G., et al., 2000. Parietal and cingulate processes in central pain. A combined positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) study of an unusual case. Pain 84 (1), 77–87.
Porcelli, A.J., Delgado, M.R., 2009. Neural systems of reward processing in humans. In: Dreher, J.C., Tremblay, L. (Eds.), Handbook of Reward and Decision Making. Academic Press, Oxford, pp. 165–184.
Postuma, R.B., Dagher, A., 2006. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. Cereb. Cortex 16 (10), 1508–1521.
Price, D.D., 2000. Psychological and neural mechanisms of the affective dimension of pain. Science 288 (5472), 1769–1772.
Redgrave, P., Gurney, K., 2006. The short-latency dopamine signal: a role in discovering novel actions? Nat. Rev. Neurosci. 7 (12), 967–975.
Rescorla, R.A., Wagner, A.R., 1972. A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement. Classical Conditioning. II. Current Research and Theory. Appleton-Century-Crofts, New York, pp. 64–66.
Roesch, M.R., Calu, D.J., Esber, G.R., Schoenbaum, G., 2010. All that glitters . . . dissociating attention and outcome expectancy from prediction errors signals. J. Neurophysiol. 104, 587–595.
Salas, R., Baldwin, P., de Biasi, M., Montague, P.R., 2010. BOLD responses to negative reward prediction errors in human habenula. Front. Hum. Neurosci. 4, 36.
Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., Nichols, T.E., 2009. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage 45 (3), 810–823.
Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599.
Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12 (3), 154–167.
Singer, T., Critchley, H.D., Preuschoff, K., 2009. A common role of insula in feelings, empathy and uncertainty. Trends Cogn. Sci. 13 (8), 334–340.
Suri, R.E., Schultz, W., 1998. Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Exp. Brain Res. 121, 350–354.
Suri, R.E., Schultz, W., 1999. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 91, 871–890.
Sutton, R.S., Barto, A.G., 1987. A temporal-difference model of classical conditioning. In: Proceedings of the Ninth Annual Conference of the Cognitive Science Society, Erlbaum, Hillsdale, New Jersey.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning. Cambridge, MIT Press.
Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16 (3), 765–780.
Vogt, B.A., 2005. Pain and emotion interactions in subregions of the cingulate gyrus. Nat. Rev. Neurosci. 6 (7), 533–544.
Voorn, P., Vanderschuren, L.J.M.J., Groenewegen, H.J., Robbins, T.W., Pennartz, C.M.A., 2004. Putting a spin on the dorsal–ventral divide of the striatum. Trends Neurosci. 27 (8), 468–474.
Yin, H.H., Knowlton, B.J., 2006. The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. 7 (6), 464–476.
van der Meer, M.A., Redish, A.D., 2011. Ventral striatum: a critical look at models of learning and evaluation. Curr. Opin. Neurobiol. 21 (3), 387–392.
van der Meer, M.A., Redish, A.D., 2010. Expectancies in decision making, reinforcement learning, and ventral striatum. Front. Neurosci. 4 (1), 29–37.