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

[1] Present address: Department of Psychology, and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK

* Corresponding author at: School of Psychology, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK. Tel: +44 1707 284638; Fax +44 1707 285073.

Email address: d.j.done@herts.ac.uk

Conflict of Interest: All authors report no conflict of interest.

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.

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

Indexing Phrase: Garrison, J.R., B. Erdeniz and D.J. Done. Prediction Error in Reinforcement Learning: A Meta-analysis of Neuroimaging studies. NEUROSCIENCE BIOBEHAV REV XX(X) XXX-XXX, 2012.

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 & Mosteller, 1951; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & 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:

$$\delta = R_t - V_{(t)} \qquad \text{[Eq. 1]}$$

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 equates to zero then learning does not occur, even when there continues to be a joint occurrence of conditional and unconditional stimuli (Schultz & Dickinson, 2000, Niv & 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:

$$\delta_t = r_t + \gamma\, V(S_{t+1}) - V(S_t) \qquad \text{[Eq.2]}$$

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 (see Sutton & Barto, 1998, for a review of SARSA and Q-learning algorithms). 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 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 reward prediction error 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.

A number of studies which explored the application of computational models , derived from machine reinforcement learning, have focused solely on the basal ganglia, and the neural circuitry that links this subcortical region to numerous 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 maximize future reward together with a critic module, which calculates a TD prediction error (Barto, 1995).

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 & 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, derived from a machine learning algorithm, has its detractors. 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. 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). Redgrave and Gurney (2006) also raise doubts about a simple interpretation of phasic DA coding reward prediction error. From a biologically informed perspective, they argue that the role of DA in reinforcement learning has more to do with reinforcing salient features of the context and the instrumental actions that are causally related to reward, and so the critic role may not be localized in one brain region but distributed.

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 organization 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 organization 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 (www.scopus.com) and Pubmed (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 derived from either of 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 excluded where (1) analysis was based solely on one or more 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 matched control group were included if coordinate data was available).

Email contact was made with the authors of 16 papers where the study met the inclusion criteria but where whole brain prediction error coordinates were not included in the original articles. This added another 7 studies making a total of 35 whole-brain activation studies (Appendix A).

2.2 Study categorisation and extraction of coordinate data

Studies were categorised according to whether the experimental task involved instrumental (Inst) or Pavlovian (Pav, classical) 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 (www.brainmap.org/ale/). RPE and APE prediction error foci represented by both positive and negative BOLD signals were extracted from 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 (Pavlovian-punishment) was 71, exceeding the minimum number of coordinates recommended for use of the GingerALE meta-analysis for a simple sensory task of 20 (www.brainmap.org/ale/). A summary of the studies used and their allocations to

each meta-analysis group is shown in Table 1 with the studies included in the meta-analysis listed in Appendix A.

The distinction between dorsal and ventral striatum is given particular consideration here due to their importance 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 & 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 & Dagher (2006) i.e.

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

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

--- Table 1 about here ---

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):

left hemisphere: $-15 < x < -3$; $4 < y < 14$; $-14 < z < -3$

right hemisphere: $14 < x < 4$; $5 < y < 15$; $-14 < z < -3$

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 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 reported 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 (e.g. instrumental or Pavlovian learning) for each study. Given the adjustment for sample size, studies with larger numbers of subjects will 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 and then doing the same 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 $50mm^3$ 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 $50mm^3$ 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 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 Figure 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 activations are listed separately (e.g. left hand column in Table 2) and the cluster size (right hand column) is left as blank.

--- Table 2 about here ---

--- Figure 1 about here ---

In the analysis of all reinforcement learning studies (Table 2 and Figure 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 about here ---

--- Figure 2 about here ---

Table 3 and Figure 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 (1-2mm at certain points) with the 1,200mm^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 findings of laterality and non-inclusion of NAcc within the Pavlovian group of studies were confirmed in a separate ALE meta-analysis of 9 further Pavlovian ROI studies (39 foci). A similar effect of laterality was observed in the right hemisphere with the Pavlovian striatal cluster located more dorsal than the cluster in the left hemisphere. The Pavlovian striatal clusters in both hemispheres lie adjacent to, but not encompassing the NAcc.

The overlap analysis shown in Figure 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 of the instrumental or Pavlovian conditions.

--- Table 4 about here ---

--- Figure 3 about here ---

Table 4 and Figure 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.

3.3 Reward and Punishment Studies

--- Table 5 about here ---

--- Figure 4 about here ---

Table 5 and Figure 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 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 reward PE 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 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 (reward) and fusiform and temporal gyri, and superior parietal lobule (punishment). 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.

--- Table 6 about here ---

--- Figure 5 about here ---

The ALE subtraction analyses of reward and punishment studies shown in Table 6 and Figure 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.

3.4 Summary of Results

This meta-analysis examined activation of human brain that correlated with reward and aversive prediction errors. In the studies included in the current meta-analysis, prediction errors were measured in several ways, but all assume an underlying model that approximates either a Rescorla–Wagner or TD prediction error, or a model derived from either of these.

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 Figure 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 & 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 ventral striatum 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 the reward and aversive prediction errors 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 & Rolls, 2011).

4. Discussion

Our findings are largely consistent with the review of the PE animal literature undertaken by Schultz & Dickinson (2000) which indicated that most areas of the brain calculate a prediction error, and which suggested that PE might be a common form of neuronal processing that is not confined to updating the expected value of the conditioned stimulus in reinforcement learning (see Fiorillo, 2008; Friston 2009; 2010; Glimcher, 2011 for similar suggestions). Therefore some caution is required since significant correlation between BOLD activity and the RPE might be the result of different cognitive processes that are themselves correlated with RPE (Roesch et al 2010). By way of example, Roesch et al., (2010) used several criteria to investigate whether neuronal activity in the amygdala which correlates with unexpected reward was actually RPE, outcome expectancy or, as they concluded, shift of attention. In response to this issue, Caplin and Dean (2008a; 2008b) and Rutledge et al (2010) suggest three axioms that need to be obeyed if the BOLD signal in a particular region meets the necessary and sufficient conditions for coding RPE (that the size of the PE signal is consistent with prize ordering, uncertainty ordering and reflects surprise equivalence).

In comparison with the ALE meta-analysis study of Liu et al. (2010) which mapped objective reward and punishment coding rather than prediction error, we note the absence of PE activation in the reward regions of medial orbitofrontal cortex, hippocampus and amygdala.

4.1 RPE in Instrumental vs. Pavlovian Reinforcement Learning

One version of the actor-critic model of reinforcement learning assumes a single critic that 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 & Redish, 2011). On this basis, 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 which 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 & 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; Baldassare 2002) or 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), or alternatively that different regions in the striatum may each contribute to a 'distributed' critic function, and the nature of this distribution is task dependent, 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 & 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).

However, the actor-critic model has many detractors as an accurate model for the functional organisation of the basal ganglia. Humphries & Prescott (2010) consider that the conflux of information channels in the NAcc coding personal goals, spatial properties of the environment as well as prediction error suggest that this area is also involved in action selection, a view endorsed by Nicola (2007). Yin & Knowlton (2006) and Nicola (2007) also consider the ventral striatum to be involved in Pavlovian-instrumental transfer.

An alternative view offered by Berridge, (2007; 2012) suggests that the ventral striatum is involved with incentive salience, or 'wanting'. A key postulate of incentive salience is that the current motivational value of the US transfers to the CS in reinforcement learning, and as such the value of the CS will vary dynamically according to motivational state. The ventral striatum is seen as critical in this aspect of incentive salience, indeed more so than for the distinct process of 'cognitive wanting' (which comprises two separate processes of memory retrieval of CS value, and the computation of prediction error). Ventral striatal activity is thus expected to depend on the motivational value of both the CS and US, and if these are low, as might be the case for some experimental fMRI study designs, then low levels of ventral striatal activity would result. This could explain 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 only occur if the agent is motivated to act).

4.2 RPE coding for reward and punishment

Our results suggest possible differences in the processing of aversive and appetitive reinforcers. 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 & Schultz, 1998; Bayer & 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 of RPE in the ventral striatum is unsurprising given neuroanatomical evidence that shows dense connections between the major axonal pathways of dopaminergic neurons and

the striatum (Arbuthnott & Wickens 2007), and fits well with the findings of a previous study (Delgado et al., 2008). In contrast, the anatomical source of the aversive PE in the left ventral striatum is less clear. Earlier literature is mixed on whether DA neurons in the midbrain code solely for rewards or whether separate midbrain DA neurons code for reward and punishment respectively (see i.e. Schultz, 2010; Bromberg-Martin et al., 2010 for a discussion).

Robust aversive PE activation clusters were also detected in both the habenula and insular cortex. The habenula finding is consistent with those of Salas et al (2010) who showed that the habenula complex calculates negative prediction error. 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). It is suggested that insular cortex might support different levels of representation of current and predictive states allowing for error-based learning relating to both feeling and uncertainty (Singer et al., 2009).

Both reward and aversive prediction error clusters were found in cingulate cortex, albeit with some regional differences within this area. Pregenual ACC was active only for reward prediction errors whereas antero-medial ACC activity correlated with both reward and aversive prediction errors. Both RPE and APE were observed in anterior cingulate cortex (ACC) in keeping with the results of electrophysiology studies in primates (Amiez et al, 2006; Matsumoto & Hikosaka, 2007), as well as EEG studies in humans (Holroyd & Coles, 2002; Nieuwenhuis et al., 2005; Oliveira et al, 2007). The antero-medial ACC has previously been reported to play a role in negative affectivity (Price, 2000) and pain (Farrell, 2005; Lancaster et al., 2000;

Peyron et al., 2000; Vogt, 2005) and be more sensitive to aversive prediction errors than reward prediction errors (Shackman et al., 2011). Our results suggest that antero-medial ACC is involved in the calculation of PE for both rewarding and punishing outcomes, but this may reflect its role in conflict-monitoring (Botvinick et al. 2004) or error-detection (Carter et al., 1998).

A final point on the analysis of reward and aversive prediction error relates to the intrinsic coupling between action and reward in the dopaminergic system. This suggests that the neural circuitry shared by reward and aversive PE in the ventral striatum might also be utilised in the processing of action requirements and motivational valence. Guitart-Masip et al., (2011) recently demonstrated the greater involvement of ventral striatum (and partially dorsal striatum) in the coding of 'go' responses (i.e. elicit more striatal activity) than 'no go' responses, a finding that was irrespective of the outcome valence (reward or punishment) . It is possible therefore that our observation of greater VS signalling of reward prediction error in instrumental learning may be confounded by DA related activity specifically relating to action ('go' as opposed to 'no go').

4.3 Caveats

This meta-analysis has several methodological caveats. Many fMRI studies of prediction error in humans have used striatal Region of Interest analysis (ROI) given the findings of electrophysiological studies in animals. In order to avoid bias in our results, all ROI studies were excluded from the meta-analysis, although we have made reference to some in order to corroborate the different findings for instrumental and Pavlovian learning.

It was also impossible to separate studies of instrumental learning into those requiring goal based responding (in computational models, these are referred to as 'model-based' or forward 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 & Prescott, 2010; Van der Meer & 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 of the meta-analytical 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 weakness of the ALE meta-analysis approach is its failure to take into account the magnitude 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.

And finally, a general caveat arises from the relation between the blood oxygenated level dependent (BOLD) signal in imaging studies and the underlying dopaminergic activity particularly in the striatum and any non-cortical structure. All of the studies included in this meta-analysis have assumed the BOLD signal used in fMRI to be an

indirect proxy of neuronal activity due to increased perfusion in response to increased glucose utilization, or increased local field potentials, both being caused by the 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, BOLD in experiments that involve DA release might be influenced by changes to perfusion resulting from both neuronal activity as well as from the direct effects of DA on vascular mechanisms.

Acknowledgements

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 Reinforcement learning theory. Cerebral Cortex, 18(7), 1485-95.

Cohen, M. X. (2007). Individual differences and the neural representations of reward expectation and reward prediction error. Social Cognitive and Affective Neuroscience, 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. Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 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. Cerebral 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. The Journal of Neuroscience, 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. Journal of Neuroscience, 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. Journal of Cognitive Neuroscience, 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 Biology, 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. Cerebral Cortex, 17(4), 749-759.

Li, J., Delgado, M. R., & Phelps, E. A. (2011a). How instructed knowledge modulates the neural systems of reward learning. Proceedings of the National Academy of Sciences of the United States of America, 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 Reinforcement learning. Nature Neuroscience, 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. Molecular 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. Molecular Psychiatry, 13, 267–276.

O'Doherty, J. P., , 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. Journal of Neuroscience, 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. Human Brain Mapping, 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. The Journal of Neuroscience, 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. Nature Neuroscience, 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. Human Brain Mapping, 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 Networks, 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. Journal of Neurophysiology, 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. Journal of Neurophysiology, 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), 261–70.

Amiez, C., Joseph, J. P., & Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. Cerebral Cortex,16(7), 1040–1055.

Arbuthnott, G.W. and Wickens, J. (2007). Space, time and dopamine. Trends in Neurosciences, 30(2), 62-69.

Baldassarre, G. (2002). A Modular Neural-Network Model of the Basal Ganglia's Role in Learning and Selecting Motor Behaviors. Journal of Cognitive Systems Research, 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 Models of Information Processing in the Basal Ganglia, J. C. Houk, J. Davis, and D. Beiser (Eds.). 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, 191(3), 391–431.

Berridge, K. C. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. The European Journal of Neuroscience, 35(7), 1124–1143.

Bornstein, A. M. & Daw, N. D. (2011). Multiplicity of control in the basal ganglia: Computational roles of striatal sub-regions. Current Opinion in Neurobiology 21(3), 374-380.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8(12), 539–546.

Bromberg-Martin, E. S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. Journal of Neurophysiology, 104(2), 1068–1076.

Bush, R. R., & Mosteller, F. A. (1951). Mathematical model for simple learning. Psychological Review, 58, 313–323.

Caplin, A, & Dean, M. (2008a). Dopamine, reward prediction error, and economics. Quarterly Journal of Economics, 123, 663–701.

Caplin, A, & Dean, M. (2008b) Axiomatic methods, dopamine and reward prediction error. Current Opinion in Neurobiology, 18, 197–202.

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.

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 Networks, 22, 213-219.

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

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363(1511), 3787–3800.

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

Doya, K., Samejima, K., Katagiri, K. & Kawato, M. (2002) Multiple Model-based Reinforcement Learning. Neural Computation. 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. Human Brain Mapping, 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. Human Brain Mapping, 25(1), 129–139.

Fiorillo, C.D. (2008). Towards a general theory of neural computation based on prediction by single neurons. PLoS One, 3: e3298.

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. Journal of Neurophysiology, 101(6), 3284–3293.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? Trends in Cognitive Sciences. 13, 293-301.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 108 (supp), 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. Journal of Neuroscience, 31, 7867–7875.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. Trends in Cognitive Sciences, 15(2), 56–67.

Haber, S. N., & McFarland, N. R. (1999). The concept of the ventral striatum in nonhuman primates. Annals of the New York Academy of Sciences, 877, 33–48.

Hikosaka, O. (2010). The habenula: from stress evasion to value-based decision-making. Nature Reviews. Neuroscience, 11(7), 503–513.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience, 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. Psychological Review, 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 J. C. Houk, J. L. Davis,

& D. G. Beiser (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. Progress in Neurobiology, 90(4), 385–417.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Networks, 15(4-6), 535–547.

Khamassi, M. (2005). Actor-Critic Models of Reinforcement Learning in the Basal Ganglia: From Natural to Artificial Rats. Adaptive Behavior, 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. Human Brain Mapping, 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. Human Brain mapping, 10(3), 120–31.

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. Neuroscience and Biobehavioral Reviews, 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. Proceedings of the National Academy of Sciences of the United States of America, 104(22), 9493–9498.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological Review, 82, 276–298.

Martin, J. J. (2003). Neuroanatomy: Text and Atlas. New York: McGraw-Hill.

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 in Neurosciences, 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. The 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. Journal of Neuroscience, 16(5),1936-1947.

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. Psychopharmacology, 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. The European Journal of Neuroscience, 21(11), 3161–3168.

Niv Y. & Schoenbaum G. (2008) - Dialogues on prediction errors - Trends in Cognitive Sciences, 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. Journal of Cognitive Neuroscience, 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. Annals of the New York Academy of Sciences. 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. Psychological Review, 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. and 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. Cerebral 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? Nature Reviews Neuroscience. 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 ll: Current research and theory (pp. 64–69). Appleton-Century-Crofts, New York.

Roesch, M. R., Calu, D. J., Esber, G. R., & Schoenbaum, G. (2010). All That Glitters . . . Dissociating Attention and Outcome Expectancy From Prediction Errors Signals. Journal of Neurophysiology, 104, 587–595.

Rutledge, R.B., Dean, M., Caplin, A., & Glimcher, P.W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. Journal of Neuroscience, 30 (40),13525-13536.

Salas, R., Baldwin, P., de Biasi, M., & Montague, P. R. (2010). BOLD Responses to Negative Reward Prediction Errors in Human Habenula. Frontiers in Human Neuroscience, 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. Annual Review of Neuroscience, 23, 473–500.

Schultz, W., Dayan, P. & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. Science, 275(5306), 1593–1599.

Schultz W. (2010) Dopamine signals for reward value and risk: basic and recent data. Behavioural and Brain Functions, 6:24.

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. Nature Reviews. Neuroscience, 12(3), 154–167.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. Trends in Cognitive Sciences, 13(8), 334–340.

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Experimental Brain Research, 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. and 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. MIT Press, Cambridge.

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. Nature Reviews. Neuroscience, 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 in Neurosciences, 27(8), 468–474.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nature Reviews. Neuroscience, 7(6), 464–476.

van der Meer, M. A., & Redish, A D. (2011). Ventral striatum: a critical look at models of learning and evaluation. Current Opinion in Neurobiology, 21(3), 387–392.

van der Meer, M. A. & Redish A. D. (2010) Expectancies in decision making, reinforcement learning, and ventral striatum. Frontiers in Neuroscience, 4(1), 29-37.

Figure Captions

Figure 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.

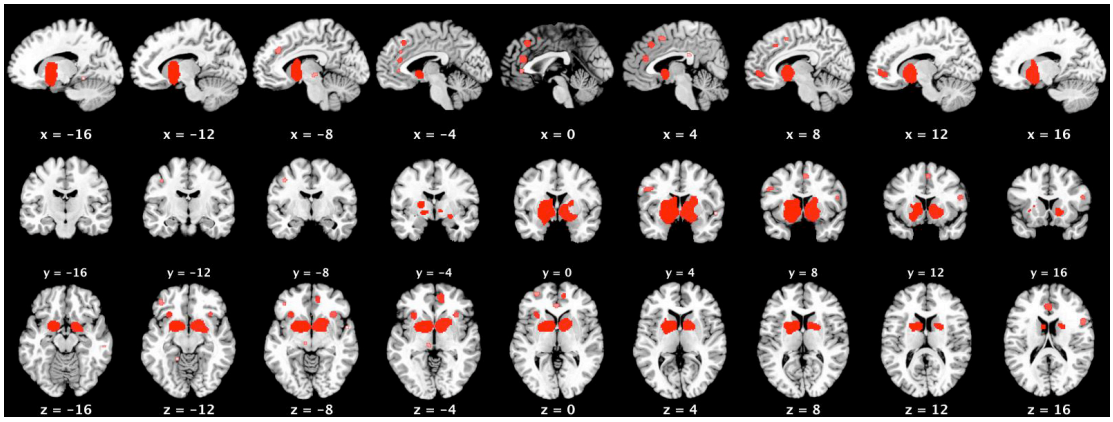

Figure 2: Results of the ALE meta-analysis for instrumental (blue) and Pavlovian (red) prediction error studies. The overlap of the two analyses is shown in green.

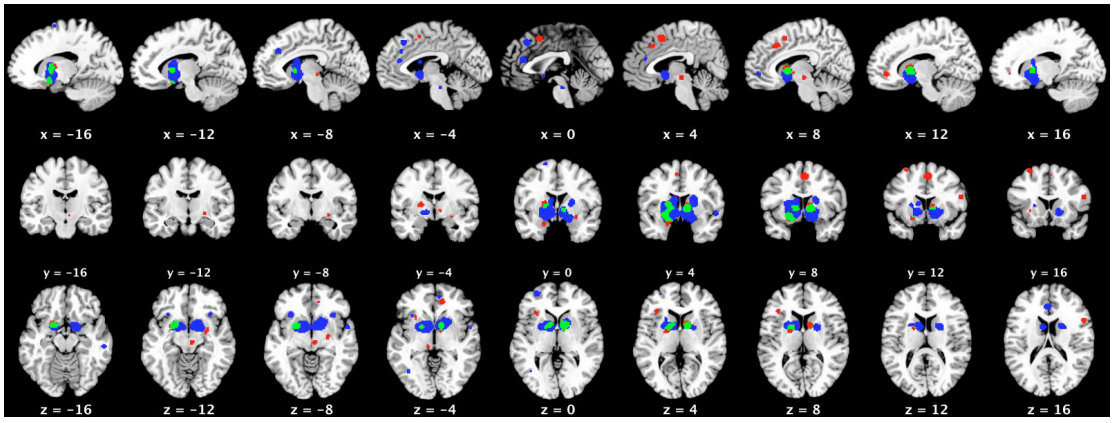

Figure 3: Results of the ALE subtraction analysis for [instrumental-Pavlovian] (blue) and [Pavlovian-instrumental] (red) prediction error studies. The overlap of the two analyses is shown in green.

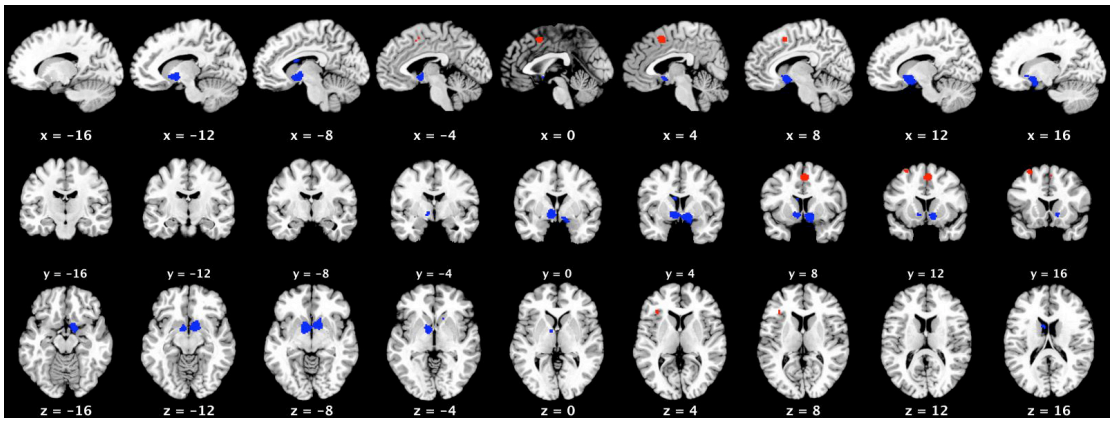

Figure 4: Results of the ALE meta-analysis for reward (blue) and aversive prediction error studies (red). The overlap of the two analyses is shown in green.

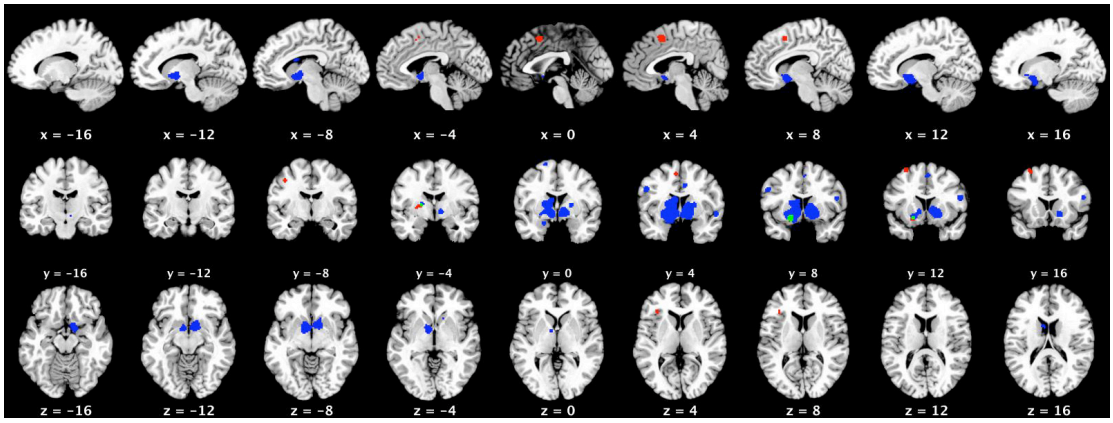

Figure 5: Results of the ALE subtraction analysis for Reward-Punishment (blue) and Punishment-Reward (red) prediction error studies. The overlap of the two analyses is shown in green.

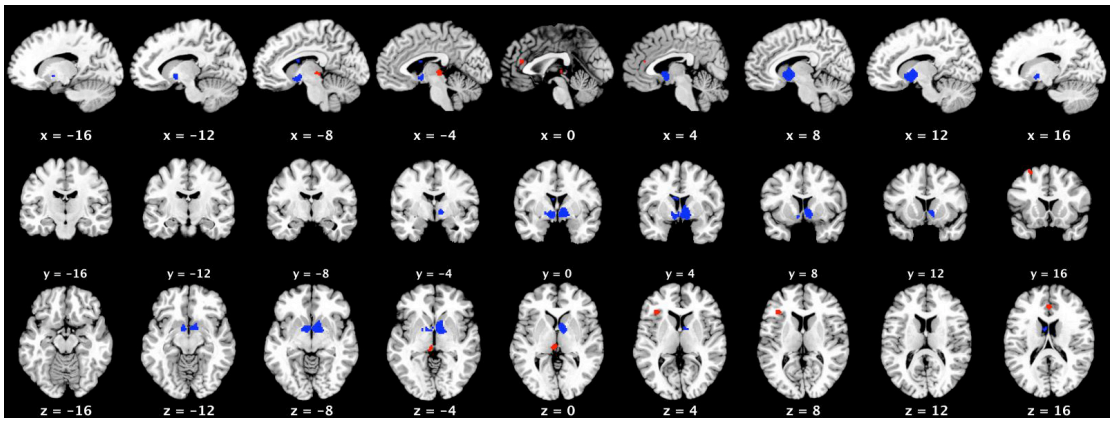

Table Captions

Table 1

Categorisation of fMRI Prediction Error Studies and Allocation to Meta-analysis Contrast Groups

Study	Learning Type	Reinforcer Type	Number of Subjects	All Studies	Inst	Pav	Reward	Punish	Combined R&P
Brovelli (2008)	Inst	Combined R&P	14	✓	✓				✓
Cohen (2007)	Inst	Reward	17	✓	✓		✓		
Delgado et al (2008)	Pav	Punishment	11	✓		✓		✓	
Gläscher (2009)	Inst	Combined R&P	20	✓	✓				✓
Gläscher (2010)	Inst	Reward	18	✓	✓		✓		
Gradin (2011)	Inst	Reward	17	✓	✓		✓		
Hampton (2006)	Inst	Combined R&P	16	✓	✓				✓
Howard-Jones (2010)	Inst	Reward	16	✓	✓		✓		
Jocham (2011)	Inst	Reward	16	✓	✓		✓		
Kahnt (2009)	Inst	Reward	19	✓	✓		✓		
Kahnt (2011)	Inst	Combined R&P	20	✓	✓				✓
Kim (2006)	Inst	Reward/Punish	16	✓	✓		✓	✓	
Kumar (2008)	Pav	Reward	18	✓		✓	✓		
Landmann (2007)	Inst	Reward	16	✓	✓		✓		
Li (2006)	Inst	Combined R&P	46	✓	✓				✓
Li (2011a)	Inst	Combined R&P	20	✓	✓				✓
Li (2011b)	Pav	Punishment	17	✓		✓		✓	
McClure (2003)	Pav	Reward	28	✓		✓	✓		
Morris (2012)	Pav	Reward	16	✓		✓	✓		
Murray (2008)	Inst	Reward	12	✓	✓		✓		
O'Sullivan (2011)	Inst	Reward	24	✓	✓		✓		
O'Doherty (2003)	Pav	Reward	13	✓		✓	✓		
Park (2010)	Inst	Combined R&P	16	✓	✓				✓
Pessiglione (2006)	Inst	Reward	13	✓	✓		✓		
Ramnani (2004)	Pav	Reward	6	✓		✓	✓		
Rodriguez (2006)	Inst	Combined R&P	15	✓	✓				✓
Schonberg (2010)	Inst	Combined R&P	17	✓	✓				✓
Seger (2010)	Inst	Combined R&P	10	✓	✓				✓
Seymour (2005)	Pav	Reward/Punish	19	✓		✓	✓	✓	
Seymour (2007)	Pav	Reward/Punish	24	✓		✓	✓	✓	
Spoormaker (2011a)	Pav	Punishment	35	✓		✓		✓	
Spoormaker (2011b)	Pav	Punishment	18	✓		✓		✓	
Tanaka (2006)	Inst	Combined R&P	18	✓	✓				✓
Tobler (2006)	Pav	Reward	22	✓		✓	✓		
Valentin (2009)	Inst	Reward	17	✓	✓		✓		
		Number of Studies:		35	23	12	20	7	11
		Number of foci:		446	293	153	262	71	120

Footnote to Table 1: For simplicity, meta-analytical studies are referred to by the name of the first author and year of publication.

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 NAc
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 NAc
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	

Footnote to Table 2: In this and all subsequent tables the following abbreviations are used:

DS: dorsal striatum, VS: ventral striatum, NAcc: Nucleus accumbens, L: left hemisphere; R: right hemisphere

Table 3

Detailed ALE Cluster Results for Instrumental and Pavlovian Prediction Error Studies

Instrumental	L/R	x	y	z	ALE (10^{-3})	Size	
Claustrum	L	-10	6	-6	49	16,528	Cluster encompasses NAc
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	L/R	x	y	z	ALE (10^{-3})	Size	
Putamen (DS)	L	-18	2	8	19	3,120	Cluster borders but does not encompass NAc.
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	1,592	Cluster extends into DS; borders but does not encompass NAc.
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	

Table 4

Detailed ALE Cluster Results for Instrumental and Pavlovian Prediction Error Studies: Subtraction Analyses

Instrumental - Pavlovian	L/R	x	y	z	ALE (10^{-3})	Size	
Caudate (VS)	R	8	10	-14	2,948	1,872	Cluster encompasses NAc
Caudate (VS)	R	10	14	-10	2,820		
Pallidum	L	-8	0	-10	2,748	1,472	Cluster extends into VS including NAc.
Caudate (DS)	L	-8	4	18	1,943	88	

Pavlovian - Instrumental	L/R	x	y	z	ALE (10^{-3})	Size
Cingulate Gyrus	R	2	12	47	3,156	840
Medial Frontal Gyrus	R	0	8	52	2,989	
Middle Frontal Gyrus	L	-28	20	56	1,999	296
Middle Frontal Gyrus	L	-30	16	56	1,960	
Insula	L	-34	27	5	1,957	120

Table 5

Detailed ALE Cluster Results for Reward and Punishment Prediction Error Studies

Reward	R	x	y	z	ALE (10^{-3})	Size	
Claustrum	L	-10	6	-6	35	7,440	Cluster encompasses NAc
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	5,360	Cluster encompasses NAc
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	R	x	y	z	ALE (10^{-3})	Size	
Putamen (VS)	L	-18	8	-14	13	400	VS cluster bordering but not including NAc.
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 DS 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	

Table 6

Detailed ALE Cluster Results for Reward and Punishment Prediction Error Studies: Subtraction Analyses

Reward - Punishment	R	x	y	z	ALE (10^{-3})	Size	
Hypothalamus	R	10	-3	-7	1,957	2,424	Cluster encompasses DS and VS including NAc
Anterior Cingulate	R	4	5	-11	1,957		
Pallidum	R	10	2	-6	1,957		
Thalamus	R	10	-4	-2	1,957		
Pallidum	R	12	0	-2	1,957		
Pallidum	R	12	2	-12	1,957		
Anterior Cingulate	L	-2	2	-9	1,957	776	Cluster encompasses VS including NAc, bordering DS at z < 1.
Pallidum	L	-7	-2	-8	1,957		
Pallidum	L	-18	0	-6	1,957		
Caudate	L	-5	1	16	1,957	88	DS cluster
Caudate	L	-10	3	18	1,957		

Punishment - Reward	R	x	y	z	ALE (10^{-3})	Size
Thalamus	L	-3	-25	1	2,297	296
Thalamus	L	-8	-29	-2	1,861	
Insula	L	-38	26	8	2,050	296
Insula	L	-34	28	8	1,977	
Anterior Cingulate	R	1	36	15	1,765	128
Middle Frontal Gyrus	L	-28	16	56	1,789	80
Middle Frontal Gyrus	L	-30	15	60	1,719	

Supplementary Table- Type of Prediction Error in Meta-Analytical Studies

	Learning Type	Type of Prediction Error
Brovelli (2008)	Inst	RW
Cohen (2007)	Inst	RW
Delgado et al(2008)	Inst	TD
Gläscher (2010)	Inst	SARSA
Gläscher (2009)	Inst	RW
Gradin (2011)	Inst	SARSA
Hampton (2006)	Inst	RW
Howard-Jones (2010)	Inst	RW
Jocham (2011)	Inst	RW
Kahnt (2011)	Inst	RW
Kahnt (2009)	Inst	RW
Kim (2006)	Inst	TD
Kumar (2008)	Inst	TD
Landmann (2007)	Inst	RW
Li (2011a)	Inst	RW
Li (2006)	Inst	RW
Li (2011b)	Inst	RW
McClure (2003)	Inst	N/A
Morris (2011)	Pav	N/A
Murray (2008)	Inst	RW
O'Doherty (2003)	Pav	TD
O'Sullivan (2011)	Inst	RW
Park (2010)	Inst	RW
Pessiglione (2006)	Inst	RW
Ramnani (2004)	Pav	N/A
Rodriguez (2006)	Inst	RW
Schonberg (2010)	Inst	TD
Seger (2010)	Inst	SARSA
Seymour (2007)	Pav	TD
Seymour (2005)	Pav	TD
Spoormaker (2011a)	Pav	TD
Spoormaker (2011b)	Pav	TD
Tanaka (2006)	Inst	SARSA
Tobler (2006)	Pav	TD
Valentin (2009)	Inst	RW

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 o
The algorithms that use TD used either the actor-critic or advantage learning rules.