Cogn Affect Behav Neurosci (2015) 15:435–459
DOI 10.3758/s13415-015-0338-7

Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis

Henry W. Chase · Poornima Kumar · Simon B. Eickhoff · Alexandre Y. Dombrovski

Published online: 10 February 2015

Abstract *Reinforcement learning* describes motivated behavior in terms of two abstract signals. The representation of discrepancies between expected and actual rewards/punishments—*prediction error*—is thought to update the *expected value* of actions and predictive stimuli. Electrophysiological and lesion studies have suggested that mesostriatal prediction error signals control behavior through synaptic modification of cortico-striato-thalamic networks. Signals in the ventromedial prefrontal and orbitofrontal cortex are implicated in representing expected value. To obtain unbiased maps of these representations in the human brain, we performed a meta-analysis of functional magnetic resonance imaging studies that had employed algorithmic reinforcement learning models across a variety of experimental paradigms. We found that the ventral striatum (medial and lateral) and midbrain/thalamus represented reward prediction errors, consistent with animal studies. Prediction error signals were also seen in the frontal operculum/insula, particularly for social rewards. In Pavlovian studies, striatal prediction error signals extended into the amygdala, whereas instrumental tasks engaged the caudate. Prediction error maps were sensitive to the model-fitting procedure (fixed or individually estimated) and to the extent of spatial smoothing. A correlate of expected value was found in a posterior region of the ventromedial prefrontal cortex, caudal and medial to the orbitofrontal regions identified in animal studies. These findings highlight a reproducible motif of reinforcement learning in the cortico-striatal loops and identify methodological dimensions that may influence the reproducibility of activation patterns across studies.

Keywords Prediction error · Expected value · Reinforcement learning · Meta analysis

Henry W. Chase and Poornima Kumar contributed equally to this work.

H. W. Chase (✉) · A. Y. Dombrovski
Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
e-mail: chaseh@upmc.edu

P. Kumar
Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, MA, USA

P. Kumar
Department of Psychiatry, Harvard Medical School, Cambridge, MA, USA

S. B. Eickhoff
Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany

S. B. Eickhoff
Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

Introduction

Behavior can be controlled by reward or punishment, and by the environmental stimuli that predict them. The way that animals develop representations of these predictive relationships has been described in terms of mathematical models of reinforcement learning, a restricted set of which have dominated experimental and theoretical attention. With the advent of new neurophysiological and imaging methods, insights from these models have advanced our understanding of the role of cortico-striato-thalamic networks, the midbrain, the amygdala, and the monoamine systems in behavioral adaptation. In particular, the activity of dopamine neurons in the mesostriatal pathway has been shown to conform to the predictions derived from formal learning rules (Waelti, Dickinson, & Schultz, 2001), and may also distinguish between particular instantiations of reinforcement learning

models (Roesch, Calu, & Schoenbaum, 2007). Combined with imaging and neurophysiology, they have helped us understand better the types of computations that take place in the reward system and the alterations observed in neurological and psychological disorders, including Parkinson's disease (M. J. Frank, 2005), depression (Kumar et al., 2008), schizophrenia (Gradin et al., 2011), eating disorders (G. K. Frank, Reynolds, Shott, & O'Reilly, 2011), addiction (Chiu, Lohrenz, & Montague, 2008), and suicidal behavior (Dombrovski, Szanto, Clark, Reynolds, & Siegle, 2013). Here, we provide an introduction to the constructs of *prediction error*—the discrepancy between the expected and obtained outcomes—and *expected value*. We then offer a brief overview of the putative neural substrates of these computations and present a meta-analysis of functional imaging studies that have examined the neural correlates of the prediction error and expected value constructs derived from reinforcement learning models.

The Rescorla–Wagner model of Pavlovian conditioning

Building on the earlier Bush–Mosteller model (Bush & Mosteller, 1951, 1953), Rescorla and Wagner (RW) developed their influential model of Pavlovian conditioning (Rescorla & Wagner, 1972). The RW model provides an account of animal learning from multiple conditioned stimuli (CSs). One challenge here is posed by the interactions between stimuli—such as the Kamin blocking effect, or diminished conditioned responding to stimulus X following AX → unconditioned stimulus (US) pairing preceded by A → US (Kamin, 1968). The dependent variable in the RW model is the unobserved, but theoretically plausible *associative strength (V) of the CS–US pairing*. Associative strength is conceptually close to the expected *reward value* of a given stimulus (at least when a single appetitive US is presented). Another innovation, which has enabled an elegant explanation of the Kamin blocking effect, was to combine the associative strength of all stimuli present on a given trial, in order to generate a *prediction error* (PE). In other words, according to RW, an outcome is surprising only to the extent that it is not predicted by any of the stimuli. Here is how the model describes the change in the associative strengths of the two stimuli after a trial in which the stimulus compound AX is followed by a US:

$$\begin{aligned} \Delta V_{\mathrm{A}} &= \alpha_{\mathrm{A}}\beta_{\mathrm{US}}(\lambda_{\mathrm{US}} - V_{\mathrm{AX}}), \\ \Delta V_{\mathrm{X}} &= \alpha_{\mathrm{X}}\beta_{\mathrm{US}}(\lambda_{\mathrm{US}} - V_{\mathrm{AX}}), \end{aligned} \tag{1}$$

where α is the learning rate for each stimulus, β is the learning rate for the US, λ_{US} is the asymptote of associative strength that the US will support, and $V_{\mathrm{AX}} = V_{\mathrm{A}} + V_{\mathrm{X}}$. Thus, if stimulus A is pretrained to the asymptote, subsequent training with the AX compound generates no PE for X. Besides blocking and overshadowing, the RW model has successfully accounted for a variety of Pavlovian and instrumental phenomena, despite a number of limitations (see Miller, Barnet, & Grahame, 1995).

Temporal difference models

Temporal difference (TD) models of animal learning, like RW, learn from PEs (Sutton & Barto, 1998), and describe an approach modeling prediction and optimal control. TD aims to predict all future rewards, discounting them over time:

$$\begin{aligned} R(t) = r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + \ldots \\ + \gamma^k r(t+k+1), \end{aligned} \tag{2}$$

where r is future reward and γ is the temporal discount factor, reflecting a preference for immediate over delayed rewards. Instead of waiting until all of the outcomes are experienced, TD estimates future rewards by repeating the following algorithm in each learning episode (time step):

$$V(t) \leftarrow V(t) + \alpha[r(t+1) + \gamma V(t+1) - V(t)], \tag{3}$$

where $\alpha[r(t+1) + \gamma V(t+1) - V(t)]$ is the prediction or *temporal difference error*, and $\gamma V(t+1)$ takes the place of the remaining terms $\gamma r(t+2) + \gamma^2 r(t+3) + \ldots + \gamma^k r(t+k+1)$.

To deal with the temporal distribution of predictive cues or response options, TD methods introduce the idea of *eligibility traces*. That is, only closely preceding (*eligible*) cues or actions are credited for reward or blamed for punishment.

TD provides a real-time account of learning that RW and other trial-level models do not. A key area of divergence between RW and TD is that TD treats rewards themselves and the cues that predict them as, in principle, equivalent, insofar as they are both stimuli that can invoke changes in the valuation of future rewards. Both conditioned cues and outcomes can influence value prediction and can elicit PEs. This innovation provides an effective account of the learning of sequences of stimuli, since conditioned cues can come to operate as reinforcers in their own right (Dayan & Walton, 2012). Moreover, the reinforcement value is collapsed into a single, common currency across different reinforcers. On the other hand, RW is a model that describes the extent to which the US (e.g., reward or punishment) can be predicted by environment stimuli. Thus the major focus of RW is the processing of the US, PEs occur only at the US, and all conditioned cues are treated as distinct entities competing to predict the US (Rescorla & Wagner, 1972). At the same time, one can see the parallel between the summed associative strengths of all presented CSs in RW and value in TD.

These differences between trial-level models such as RW and TD lead to differential predictions regarding the putative neural learning signals, as is illustrated in Fig. 1. A trial-level model aligns its associative strength (or expected value) signal with the CS, and PE with the US. One can see that, when the signals from a trial-level model such as RW are aligned with stimuli in real time, the time course of TD error approximates the combination of associative strength at the CS and PE at the US. On the other hand, in trial-by-trial functional magnetic resonance imaging (fMRI) learning experiments with short and, especially, fixed CS–US intervals, the predicted blood oxygenation level dependent (BOLD) signal corresponding to the associative strength or value generated by trial-level models will often approximate those of TD.

Neural correlates of prediction errors: model-based neuroimaging and electrophysiology

Prediction-error-based learning models have also enabled neuroscientists to interpret neural signals, most prominently from midbrain dopaminergic neurons (Schultz, Dayan, & Montague, 1997). The firing rates in dopaminergic neurons in this region are consistent with the predictions of RW: A blocking experiment revealed that firing rates reflect the contingency between a stimulus and a reward, rather than the mere pairing of the two (Waelti et al., 2001). Moreover, specific predictions of the TD model were also corroborated in these neurons: Most notably, neural firing within dopaminergic neurons in the midbrain gradually becomes coupled to predictive stimuli rather than to the rewards themselves (Schultz et al., 1997). In addition, a study of conditioned inhibition revealed that an inhibitory cue, predictive of reward omission, could reduce the firing rates of subpopulations of these neurons (Tobler, Dickinson, & Schultz, 2003).

A natural development of this work was to apply the same behavioral paradigms and reasoning to human neurophysiological research. Although event-related potential and magnetoencephalographic research has attempted to address analogous questions (Holroyd & Coles, 2008; Krigolson, Hassall, & Handy, 2014), the relatively limited capability of these methods to register unambiguous physiological responses from subcortical or brainstem regions has meant that the majority of progress must depend on fMRI. Since one of the seminal studies of this field (O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003), the primary focus of fMRI studies has generally been the ventral striatum, rather than the midbrain itself. A typical explanation (see, e.g., Roesch, Calu, Esber, & Schoenbaum, 2010; Tobler, O'Doherty, Dolan, & Schultz, 2006) is that the fMRI response reflects the phasic input to a structure (Logothetis & Pfeuffer, 2004), rather than the local processing or the region's output. Thus, given that the dopaminergic neurons of the ventral tegmental area (VTA)

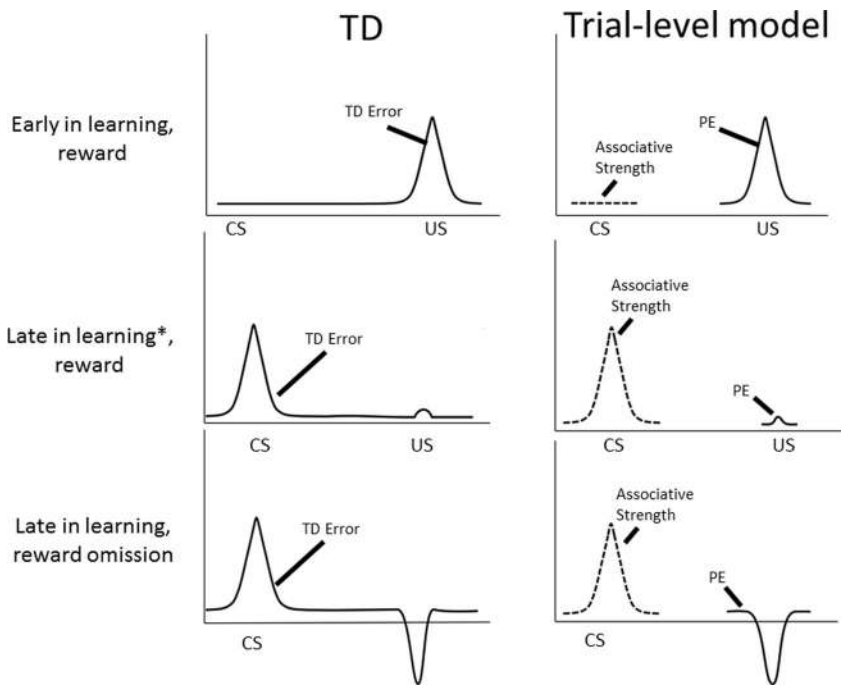

Fig. 1 The temporal difference (TD) model describes a real-time course of reward prediction error (PE) signals; PEs transfer from the unconditioned stimulus (US) to the conditioned stimulus (CS) as learning progresses. In contrast, trial-level models such as Rescorla–Wagner describe PE only at the US, whereas associative strength (conceptually close to value) signals build at the CS. It is easy to see the resemblance between the TD error signal and the combination of PE and associative strength signals in trial-level models. *Before the asymptote is reached. At asymptote, PE at the US disappears

project to the areas of the striatum (Haber, Fudge, & McFarland, 2000), fMRI-measured ventral striatal activation might then be seen as the downstream consequence of VTA firing. This perspective has found considerable support in the literature, although there are two areas of possible complication. First, there is evidence of prediction-error-related activation in the VTA itself (e.g., D'Ardenne, McClure, Nystrom, & Cohen, 2008), implying that local processing may also be relevant. Second, the ventral striatum also receives input from a wide range of cortical and subcortical regions (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2004), any of which could influence its activity and information processing within it. A further advantage of fMRI is that, although focused analysis of PE responses in the VTA and ventral striatum has been performed with this technique (D'Ardenne et al., 2008), its capability to identify signal across the entire brain has allowed for an examination of related signals in other parts of the cortex. Integration and analysis of the rich data sets obtained using fMRI methods are the focus of the present work.

Learned value, economic subjective value, and their neural correlates

In economics, *subjective value* or *utility* is the theoretical common currency used to compare disparate goods. Economic commodities can be thought of as reinforcers, and labor or a price paid as analogues of effort during operant conditioning (Lea, 1978). Although economic decision-making has traditionally been studied using stylized description-based prospects, recent research has suggested that experience-based experiments resembling animal-learning paradigms provide complementary models of real-life economic decision-making (Hertwig & Erev, 2009). Thus, to the degree that economic preferences incorporate one's reinforcement history, one may hypothesize that revealed preferences and feedback-based animal learning depend on similar neural computations (Fellows, 2011). One of the motivations for the present analysis was to examine whether the cortical regions tracking learned reward value coincide with the medial prefrontal regions that have been shown to signal economic subjective value on revealed preference tasks (Peters & Buchel, 2010).

In addition, animal electrophysiological studies have shown responses that accord well with what might be expected of learned-value signals in regions including the ventral prefrontal cortex (vPFC) and limbic areas such as the cingulate, and the striatum (Samejima, Ueda, Doya, & Kimura, 2005; Simmons, Ravel, Shidara, & Richmond, 2007; Wallis & Miller, 2003). Here, the vPFC refers to the orbitofrontal cortex (OFC), the ventromedial prefrontal cortex (vmPFC), and more lateral regions of the ventral prefrontal cortex. The vmPFC denotes the mammalian paralimbic agranular/dysgranular prefrontal cortex, encompassing monkey areas 14, 25, and rostral 24 and 32 of Petrides and Pandya (1994), and human areas 25 and rostral 32 and 24; the orbital aspect of this region is also referred to as the *medial orbitofrontal cortex* (mOFC). Associative signals represented in the vPFC possess many properties of abstract value, in that they are sensitive to delays and probability of reward, as well as to the presence of alternatives (Kennerley, Dahmubed, Lara, & Wallis, 2009; Kennerley & Wallis, 2009b; Kobayashi, Pinto de Carvalho, & Schultz, 2010; Padoa-Schioppa & Assad, 2008; Roesch & Olson, 2005; Tremblay & Schultz, 1999). These signals are "subjective," integrating such internal states as hunger (Bouret & Richmond, 2010; Critchley & Rolls, 1996). Other decision-related signals have been found in motor prefrontal and parietal cortex (Platt & Glimcher, 1999). However, it appears that these signals may reflect salience (Leathers & Olson, 2012) or motivation (Roesch & Olson, 2004), rather than value.

The present meta-analysis

The present work provides a quantitative summary of fMRI evidence on PE and expected value representations in the human brain using an activation likelihood estimation (ALE) meta-analysis. It extends recent meta-analyses of value and PE signals (Bartra, McGuire, & Kable, 2013; Clithero & Rangel, 2014; Garrison, Erdeniz, & Done, 2013; Levy & Glimcher, 2012) in two ways. First, to control methodological heterogeneity, our analysis included only studies that have used delta-rule reinforcement learning models. This enabled a better-controlled evaluation of the consequences of variations in methodology. We could thus identify the core networks that are most reliably detected. Second, to reveal the distributed networks that subserve human reward learning, we jointly mapped the regions responsive to value and PE. On the basis of the animal and human literature reviewed above, we hypothesized that PE signals would be observed in the striatum (including putamen, caudate, and nucleus accumbens) and midbrain. In contrast, we hypothesized that expected value signals would be represented in the vmPFC.

In contrast to previous meta-analyses (Bartra et al., 2013; Garrison et al., 2013; Levy & Glimcher, 2012), we focused only on studies in which signals derived from a reinforcement learning algorithm served as explanatory variables in the analysis of fMRI data. This allowed us to examine whether differences in approaches to generating such signals could yield different neural maps. We also examined other methodological variables that could have an impact on the observed coordinate maps derived from reward prediction error (RPE) experiments. Our variables of theoretical interest included instrumental or Pavlovian designs and reinforcer type (monetary, liquid, or social). Accounting for the effects of these variables would demonstrate the degree to which the RPE

maps are dependent on choices of experimental parameters. To this end, we had several secondary hypotheses.

1. *Pavlovian versus instrumental paradigms*: Prior studies had suggested differential roles for striatal subregions in Pavlovian versus instrumental tasks. Pavlovian RPEs recruit the ventral striatum, whereas RPEs from instrumental tasks (most of which include a Pavlovian component) appear to recruit both ventral and dorsal striatum (O'Doherty et al., 2004).
2. *Fixed/individual learning*: All models evaluated in the present work include a parameter that controls the rate at which conditioning occurs. There are three main strategies for determining the learning rate, all of which are evaluated in a study by Cohen (2007). He compared the neural correlates of the parameters generated by individual fits of each participant's responses ("individual") with the correlates of either the group means of such parameters ("group fixed") and an arbitrary fixed estimate of the group response ("fixed"). Despite somewhat different patterns of activation, the two methods were broadly consistent in indexing similar limbic and prefrontal regions of interest. In general, individually fitted parameters can arguably better accommodate the subject's behavior (Estes & Maddox, 2005), and thus may provide a more optimal fit of the underlying neural signals. Yet noisy, stochastic behavior, or directed exploration, may deleteriously affect the reliability of estimated parameters. Group-fitting ("group fixed") of parameters provides a form of regularization (Daw, 2011), leading to more a conservative parameterization that is potentially less susceptible to such misspecification. It may also be well suited to studies of patient groups (e.g., Bernacer et al., 2013). We tested whether each approach biased the discovery of particular brain regions. Alternatively, either approach could simply be a more accurate way of characterizing the neural correlates of individual acquisition curves, and thus be associated with similar, if more finely resolved, patterns of activation.
3. *US-aligned outcome PE versus CS- and US-aligned TD error*: As we noted above, the time course of TD error differs from that of the outcome PE generated by trial-level models. It has been suggested that TD error may be exclusively represented in the ventral striatum, whereas outcome PE is signaled by a larger network including the caudate (Niv, Edlund, Dayan, & O'Doherty, 2012). Moreover, exclusively outcome-coupled PE regressors may be more susceptible to ongoing activation coupled to the outcome, distinct from PE itself, such as the appetitive response to a rewarding outcome (Rohe, Weber, & Fliessbach, 2012). We contrasted TD and outcome PE studies, expecting to see more extensive activation to outcome PE and also anticipating that a conjunction analysis would reveal the ventral striatum as the site of overlap between these studies.
4. *Reward type*: Previous meta-analyses have examined patterns of activation in response to various primary and secondary rewards (Sescousse, Caldu, Segura, & Dreher, 2013). However, any differences and commonalities may have been driven by sensory properties of the rewarding stimuli. By contrast, our focus on model-estimated PEs allowed us to examine the spatial segregation or dissociation of more abstract neural computations triggered by disparate rewards. On the basis of the animal studies reviewed above, we hypothesized that the ventral striatum would be the shared area of activation for all types of rewards.
5. *Smoothing*: A variable without theoretical interest that might affect the pattern of data was the smoothing kernel employed by the study. Recently, Sacchet and Knutson (2013) have shown that the application of large smoothing kernels can bias the localization of ventral striatal responses to reward anticipation. In addition, it is not easy to detect BOLD activations in subcortical, and especially brainstem, nuclei because of their small size: only 60 mm^3 for the nucleus of VTA, for example (Paxinos & Huang, 1995). Yet, when preprocessing whole-brain fMRI images, researchers often use spatial filters exceeding the size of potential signal sources in these nuclei. The matched filter principle suggests that such large filters are likely to reduce the signal-to-noise (SNR) ratio in these structures. We tested whether this size mismatch affected the detection of PE signal sources in the basal ganglia and midbrain. We contrasted studies that used smaller (<8-mm) filters with those that used larger filters.

Method

Study selection criteria and definitions

Studies were selected by searching PubMed and Google Scholar to identify fMRI studies that employed computational algorithms to investigate the neural correlates of reinforcement learning studies. Combinations of keywords were used: ["reinforcement learning" OR "reward learning"], ["prediction error" OR "expected value"], and ["rescorla-wagner" OR "temporal-difference" OR "Q-learning"]. We also identified studies using reference tracing and citations within reviews. The search yielded 40 studies. Each article was reviewed by at least two authors to make sure that it fulfilled the following criteria:

1. Only studies that used a reinforcement learning model (i.e., trial-level delta-rule model, TD, or back-

propagating connectionist model) to create regressors for a general linear model (GLM) analysis of BOLD signal were included. The common feature of these studies was a PE-based learning rule.
2. Our PE analyses used maps that revealed a positive coupling with appetitive "signed" RPEs, which are positive when the reward is higher than expected or negative when it is lower than expected. Maps reporting aversive PEs were excluded, since their number was insufficient for an ALE analysis. Similarly, negative correlations with RPE or expected value (EV) regressors were also not analyzed, since these are not systematically reported.
3. EV was defined as the extent to which stimuli or actions were predictive of reward.
4. Studies that had used modified delta-rule algorithms were included as long as they involved no additional equations or components that would fundamentally change the representational structure (e.g., an upper layer in a hierarchical model).
5. Studies in which a reinforcement learning model of the sort described above was refuted or outperformed by a model from a different class (e.g., by a hidden-Markov model, Kalman filter, hierarchical Bayesian model, or hybrid models with separate representational systems) were excluded, to avoid the inclusion of maps derived from potentially disadvantaged models.
6. Only studies reporting whole-brain results were included.[1] For studies reporting only region-of-interest or otherwise restricted analyses, we contacted the authors to obtain whole-brain coordinates and included the study if the data were received.
7. We included only studies of nonclinical adult populations, excluding rare genotypes, subclinical psychopathology, and placebo-treated participants.

In total, we included in our ALE analyses 38 studies reporting RPE maps and 16 studies reporting EV maps, with 751 and 337 participants, respectively. Of the EV studies, two did not contribute RPE maps. The details of all included studies are listed in Tables 1, 2 and 3, and proportions of different study designs are displayed in Fig. 2.

[1] A study by Wittmann and colleagues (Wittmann, Daw, Seymour, & Dolan, 2008) was not included because their sequence was optimized for ventral structures, and regions above the dorsal anterior cingulate were not imaged. However, because this study could potentially have been included given alternative criteria, we compared this RPE map with those from the other studies. The RPE activations reported in this study were highly comparable with those in similarly designed (fixed, instrumental, monetary, TD) studies (e.g. putamen, visual cortex, thalamus, and opercular activation).

Subgroup analyses

Various subgroup analyses investigated heterogeneity across our studies. We classified the studies into the following categories:

- *Pavlovian/instrumental*: In "instrumental" paradigms, outcome is contingent on a behavioral response (choice). In "Pavlovian" paradigms, outcome is not contingent on choice, although a response may be made—for example, in order to signal outcome probability.
- *Fixed/individual*: A "fixed" learning rate is assumed to be equivalent for all participants within the cohort. The learning rate may be estimated at the group level (e.g., Bernacer et al., 2013) or by taking a reasonable heuristic (often around 0.2; e.g., Kumar et al., 2008). Alternatively, "individual" learning rates are estimated separately for each participant, and the PE and EV signals for each participant reflect the individually estimated learning rate.
- *Outcome PE/TD*: Although a wide variety of algorithms were used, we made a broad distinction between RW-like trial-level models and TD-like algorithms. Put simply, trial-level models have a single update mechanism at the time of the outcome that forms the basis of the RPE, whereas RPEs are computed at both the stimulus/action and outcome phases of the task in TD algorithms.
- *Monetary/liquid/cognitive/social*: "Monetary" and "liquid" paradigms involved the respective reinforcers; "cognitive" paradigms employed cognitive reinforcement, such as numerical or symbolic feedback; and "social" paradigms involved smiles, frowns, fearful, or beautiful faces as reinforcement.
- *High/low smoothing*: "High" studies employed a smoothing kernel of 8 mm or more; "low" studies employed a smoothing kernel of 7 mm or less.

Where there was a choice of maps to use from a given study that fulfilled our criteria, we selected the one in which the GLM regressor was estimated on the basis of the largest number of trials. For example, we included the overall social and monetary RPE maps reported in the study of Fareri, Chang, and Delgado (2012) for the main RPE analysis, but the social RPE map only for all of the subgrouping analyses. Other arbitrary choices included the decision to include the liquid reinforcement map in Metereau and Dreher (2013), due to the relatively low number of these studies. Finally, where slightly different models were fitted to the data, the better-fitting or otherwise preferred model was selected.

Activation likelihood estimation

Our statistical analysis of the studies was conducted using the revised activation likelihood estimation (ALE) algorithm

Table 1 Studies reporting reward prediction error (PE) maps, including details about sample size (n) and number of foci, learning rule (US = unconditioned stimulus, TD error = temporal difference error), Pavlovian/instrumental design, learning rate parameter estimation (Fixed = fixed at group level, Individual = individually estimated per participant), and reinforcer type

Study	n	Foci	Learning Rule/ PE Time Course	Pavlovian/ Instrumental	Learning Rate Parameter	Reinforcer Type
Bellebaum, Jokisch, Gizewski, Forsting, & Daum, 2012	15	52	Outcome PE	Instrumental	Individual	Monetary
Bernacer et al., 2013	18	5	Outcome PE	Instrumental	Fixed	Monetary
Bray & O'Doherty, 2007	28	6	Outcome PE	Pavlovian	Individual	Social[2]
Brovelli, Laksiri, Nazarian, Meunier, & Boussaoud, 2008	14	2	Outcome PE	Instrumental	Individual	Cognitive
Chowdhury et al., 2013	32	35	Outcome PE	Instrumental	Individual	Monetary
Dombrovski et al., 2013	20	16	Outcome PE	Instrumental	Individual	Cognitive
Fareri et al., 2012	18	6	Outcome PE	Instrumental	Individual	Monetary & Social (Social only for subgroup analysis)
Gershman, Pesaran, & Daw, 2009	16	2	Outcome PE	Instrumental	Individual	Monetary
Glascher, Hampton, & O'Doherty, 2009	20	10	Outcome PE	Instrumental	Individual	Monetary
Gradin et al., 2011	17	16	Outcome PE	Instrumental	Fixed	Liquid
Howard-Jones, Bogacz, Yoo, Leonards, & Demetriou, 2010	16	20	Outcome PE	Instrumental	Individual	Monetary
Jocham, Klein, & Ullsperger, 2011	16	13	Outcome PE	Instrumental	Individual	Monetary
Jones et al., 2011	36	12	Outcome PE	Instrumental	Fixed	Social
Kahnt et al., 2009	19	17	Outcome PE	Instrumental	Individual	Social
Kim, Shimojo, & O'Doherty, 2006	16	4	TD error	Instrumental	Individual	Monetary
Klein et al., 2007	12	4	Outcome PE	Instrumental	Individual	Social
Kumar et al., 2008	18	7	TD error	Pavlovian	Fixed	Liquid
Li, McClure, King-Casas, & Montague, 2006	46	5	Outcome PE[3]	Instrumental	Individual	Cognitive
Madlon-Kay, Pesaran, & Daw, 2013	20	8	Outcome PE	Instrumental	Individual	Monetary
Metereau & Dreher, 2013	20	20	Outcome PE	Pavlovian	Individual	Liquid[4]
Murray et al., 2008	12	17	Outcome PE	Instrumental	Fixed	Monetary
Niv et al., 2012	16	5	TD error	Instrumental	Individual	Monetary
O'Doherty et al., 2003	9	17	TD error[5]	Pavlovian	Fixed	Liquid
O'Sullivan, Szczepanowski, El-Deredy, Mason, & Bentall, 2011	24	1	Outcome PE	Instrumental	Fixed	Monetary
Park et al., 2010	16	33	Outcome PE	Instrumental	Individual	Social
Robinson, Overstreet, Charney, Vytal, & Grillon, 2013	24	7	Outcome PE	Pavlovian	Fixed	Social
Rodriguez, 2009	14	5	Outcome PE	Instrumental	Fixed	Cognitive
Rodriguez, Aron, & Poldrack, 2006	15	1	Outcome PE	Instrumental	Fixed	Cognitive
Schlagenhauf et al., 2012	28	28	Outcome PE	Instrumental	Individual	Social
Schonberg, Daw, Joel, & O'Doherty, 2007	29	14	TD error	Instrumental	Fixed	Monetary
Schonberg et al., 2010	17	22	TD error	Instrumental	Individual	Monetary
Seger, Peterson, Cincotta, Lopez-Paniagua, & Anderson, 2010	11	16	Outcome PE	Instrumental	Individual	Cognitive
Seymour et al., 2005	19	2	TD error	Pavlovian	Fixed	Relief
Takemura et al., 2011	23	8	Outcome PE[6]	Pavlovian	Fixed	Liquid
Tanaka et al., 2006	18	2	Outcome PE	Instrumental	Individual	Monetary[7]
Valentin & O'Doherty, 2009	17	37	Outcome PE	Instrumental	Fixed	Monetary & Liquid
van den Bos, Cohen, Kahnt, & Crone, 2012	22	65	Outcome PE	Instrumental	Individual	Cognitive
Watanabe, Sakagami, & Haruno, 2013	20	5	Outcome PE	Instrumental	Individual	Monetary

[2] Opposite sex – Unattractive face; [3] Matching shoulder → Rising optimum; logistic fitting map; [4] Monetary also available; [5] Results are for PE@$_{CS}$ inclusively masked with signed PE@$_{UCS}$; [6] "With" model selected, including similarity parameter; [7] "Random" condition

Table 2 Studies reporting expected value (EV) maps

Study	*n*	Foci	Pavlovian/ Instrumental	Learning Rate Parameter	Reinforcer Type
Bernacer et al., 2013	18	2	Instrumental	Fixed	Monetary
Chowdhury et al., 2013	32	100	Instrumental	Individual	Monetary
Dombrovski et al., 2013	20	4	Instrumental	Individual	Cognitive
FitzGerald, Friston, & Dolan, 2012	26	48	Instrumental	Individual	Monetary
Glascher et al., 2009	20	15	Instrumental	Individual	Monetary
Gradin et al., 2011	17	8	Instrumental	Fixed	Liquid
Jones et al., 2011	36	1	Instrumental	Fixed	Social
Kim et al., 2006	16	2	Instrumental	Individual	Monetary
Klein et al., 2007	12	8	Instrumental	Individual	Social
Madlon-Kay et al., 2013	20	6	Instrumental	Individual	Monetary
O'Sullivan et al., 2011	24	3	Instrumental	Fixed	Monetary
Seger et al., 2010	11	11	Instrumental	Individual	Cognitive
Takemura et al., 2011	23	24	Pavlovian	Fixed	Liquid
Tanaka et al., 2006	18	4	Instrumental	Individual	Monetary
Watanabe et al., 2013	20	2	Instrumental	Individual	Monetary
Wunderlich, Rangel, & O'Doherty, 2010	24	11	Instrumental	Individual	Monetary

(Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012) for coordinate-based analyses (Turkeltaub, Eden, Jones, & Zeffiro, 2002). The method generates meta-analytic maps of consistent brain activation locations from the coordinates derived from neuroimaging studies with similar experimental conditions. The method provides an estimate of the convergence of foci across activation maps, and determines the significance of these estimates via an empirically derived null distribution (Eickhoff et al., 2012). The null hypothesis is that the foci are distributed randomly across the brain, and the test statistic supports a random-effects inference, that the modeled activation maps reflect an above-chance convergence across studies (Eickhoff et al., 2012; Turkeltaub et al., 2012). A detailed description of the ALE technique can be found elsewhere (Eickhoff et al., 2012; Turkeltaub et al., 2012). In short, the activation foci reported for a given experiment are treated as centers of a 3-D Gaussian probability distribution, the width of which is empirically derived and reflects an estimate of the spatial uncertainty of the foci of a given map and the sample size of each experiment (Eickhoff et al., 2009). On the basis of the ICBM tissue probability maps, each focus is given a probability value of how likely the activation is to be located at exactly that position. One modeled activation map is then created for each experiment by merging the probability distributions of all activation foci. If more than one focus from a single experiment is jointly influencing the modeled activation map, then the maximum probability associated with any one focus reported by the given experiment is used. ALE scores are then calculated by taking the union of these individual modeled activation maps, and these scores reflect the voxel-wise convergence of activations across experiments. The p values of the ALE scores are determined with reference to the null distribution. The resulting nonparametric p values were transformed into z scores and thresholded at a cluster-level family-wise error rate-corrected threshold of $p < .05$ (cluster-forming threshold at voxel-level $p < .001$).

Comparison of the different subgroups was performed by subtracting the voxel-wise modeled activation maps from one another, and then comparing this map to an empirically derived null distribution of ALE-difference scores (10,000

Table 3 Overall numbers of participants and foci contributing to each of the contrasts investigated

	Studies	Participants	Foci
Reward PE	38	751	545
EV	16	337	249
Fixed	14	275	149
Individual	24	476	395
Instrumental	31	610	477
Pavlovian	7	141	67
RW	31	627	473
TD	7	124	71
Monetary	16	305	215
Liquid	5	87	68
Cognitive	7	142	110
Social	7	181	112

For the categories included in the subgroup analysis ("Fixed" and below), only the studies and accompanying statistics that are included in the final analyses are shown in the table

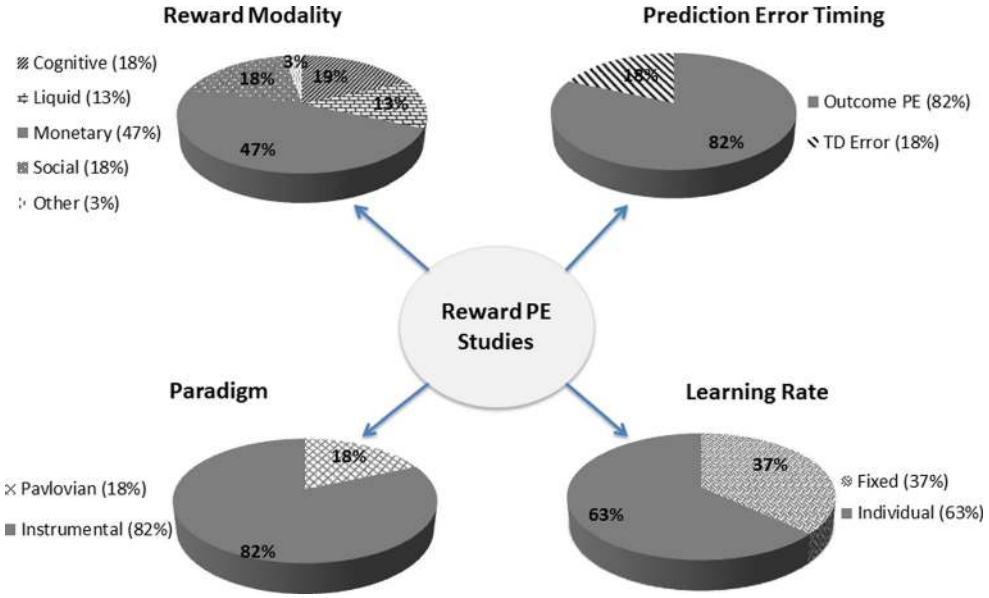

Fig. 2 Pie charts showing the percentages of studies in each condition that were included in producing the activation likelihood estimation (ALE) maps for reward prediction error

permutations). To this end, ALE analyses were performed separately on the experiments associated with either condition and the voxel-wise differences were computed between the ensuing ALE maps. All experiments contributing to either analysis were then pooled and randomly divided into two groups of the same size as the two original sets of experiments defined by activation in the first or second cluster (Eickhoff et al., 2011). The ALE scores for these two randomly assembled groups were calculated, and the difference between these ALE scores was recorded for each voxel in the brain. Repeating this process 10,000 times yielded a null distribution of differences in ALE scores between the ALE analyses of the two clusters. The "true" difference in ALE scores was then tested against this null distribution, yielding a posterior probability that the true difference was not due to random noise in an exchangeable set of labels based on the proportion of lower differences in the random exchange. The resulting probability values were then thresholded at $p > .95$ (i.e., 95 % chance for a true difference) and a cluster size (k) of 20.

Results

Reward prediction error

The activations revealed by the main categories were largely in line with our hypotheses (Table 4, Figs. 3 and 4). The ALE meta-analysis of the RPE maps revealed clusters encompassing bilateral ventral striatum, bilateral amygdala, midbrain, thalamus, frontal operculum, and insula. The largest clusters were seen in the ventral striatum: one activation cluster in each hemisphere that extended from the ventromedial caudate (nucleus accumbens) to the lateral putamen and amygdala (predominantly the superficial subregion). The left frontal operculum cluster impinged on both the pars orbitalis of the inferior frontal gyrus and the anterior insula. RPE-related activation was also observed in the left visual cortex, predominantly located in V3 and V4.

RPE: subgroup analysis

We performed a number of analyses focused on different subcategories of the RPE studies, in order to identify the distinct activations associated with different designs. First, in order to interpret these contrasts appropriately, we examined the extents to which the different categories of experimental designs were statistically independent.

Confounding

Fisher's exact tests between the subcategories assessed the contingencies between design factors. There was a highly significant association between reinforcer type and Pavlovian/instrumental design (exact test = 14.67, $p < .001$). Monetary reinforcers were more common in instrumental studies, and liquid reinforcers were more common in Pavlovian studies. Three other relationships showed trend-level associations (ps between .061 and .088): fixed/individual versus Pavlovian/instrumental, outcome PE/TD error versus reinforcer type, and outcome PE/TD error versus Pavlovian/instrumental.

This confounding between Pavlovian designs, liquid reinforcers, and TD modeling proved relevant, because the activations associated with Pavlovian designs were mostly collected from studies employing liquid

Table 4 ALE clusters representing reward prediction errors, including peak *t* statistics, Montreal Neurological Institute (MNI) coordinates, and cluster sizes

Region	*t* Statistic	Coordinates	Size	Studies Participating (Percentage Contribution)
Left striatum (ventral putamen and caudate), amygdala (SF)	6.66 5.39 3.50	–20 6 –12 –10 8 –6 –28 –6 –18	615	van den Bos et al., 2012 (10.23) Gradin et al., 2011 (8.54) Murray et al., 2008 (7.44) Bellebaum et al., 2012 (6.58) Kumar et al., 2008 (6.21) Glascher et al., 2009 (6.20) Metereau & Dreher, 2013 (5.86) Madlon-Kay et al., 2013 (5.53) Kahnt et al., 2009 (5.20) Kim et al., 2006 (4.97) Niv et al., 2012 (4.97) Seger et al., 2010 (4.94) Fareri et al., 2012 (4.85) Tanaka et al., 2006 (4.45) Howard-Jones et al., 2010 (3.22) J. P. O'Doherty et al., 2003 (2.89) Bray & O'Doherty, 2007 (2.26) Klein et al., 2007 (2.00) Seymour et al., 2005 (0.32) Jones et al., 2011 (1.97) Jocham et al., 2011 (0.21) Li et al., 2006 (0.17)
Right striatum (ventral putamen and caudate), amygdala (SF)	4.67 4.65 4.62 4.40 4.38 3.42	10 8 –10 26 –2 –12 16 8 –4 18 16 –6 14 6 –14 34 2 –12	463	Glascher et al., 2009 (8.89) Metereau & Dreher, 2013 (8.73) Kumar et al., 2008 (8.63) van den Bos et al., 2012 (7.91) Li et al., 2006 (7.79) Seger et al., 2010 (7.36) Madlon-Kay et al., 2013 (7.35) Kahnt et al., 2009 (7.23) Kim et al., 2006 (6.07) Gradin et al., 2011 (6.06) Watanabe et al., 2013 (5.77) Klein et al., 2007 (4.63) Murray et al., 2008 (3.35) Howard-Jones et al., 2010 (2.24) Fareri et al., 2012 (1.89) Jones et al., 2011 (1.62) Brovelli et al., 2008 (1.21) Schonberg et al., 2007 (1.03) J. P. O'Doherty et al., 2003 (0.78) Park et al., 2010 (0.53)
Left insula, frontal operculum	6.14	–32 24 –8	201	Jones et al., 2011 (17.89) Schlagenhauf et al., 2012 (13.23) Jocham et al., 2011 (13.00) Chowdhury et al., 2013 (12.74) Kahnt et al., 2009 (12.39) Park et al., 2010 (10.46) Seger et al., 2010 (7.19) Valentin & O'Doherty, 2009 (5.89) Glascher et al., 2009 (2.09) Robinson et al., 2013 (1.87) J. P. O'Doherty et al., 2003 (1.55) van den Bos et al., 2012 (0.26) Murray et al., 2008 (0.18)
Midbrain, thalamus	5.63	–10 –20 –6	162	Murray et al., 2008 (15.24) Bellebaum et al., 2012 (15.12) Jocham et al., 2011 (14.75) J. P. O'Doherty et al., 2003 (12.76) Rodriguez, 2009 (11.69)

Table 4 (continued)

Region	*t* Statistic	Coordinates	Size	Studies Participating (Percentage Contribution)
				Valentin & O'Doherty, 2009 (11.20) Jones et al., 2011 (9.19) Kumar et al., 2008 (4.68) Park et al., 2010 (1.63) Seymour et al., 2005 (1.21) Gradin et al., 2011 (1.15) Schlagenhauf et al., 2012 (0.44)
Left fusiform, lingual, inferior occipital gyrus (V3, V4)	4.08 4.05 3.87 3.18	–22 –82 –18 –34 –84 –8 –24 –88 –16 –24 –82 –8	147	Chowdhury et al., 2013 (23.64) van den Bos et al., 2012 (17.92) Bellebaum et al., 2012 (13.15) Schonberg et al., 2010 (11.75) Gradin et al., 2011 (9.43) Madlon-Kay et al., 2013 (8.96) Howard-Jones et al., 2010 (7.83) O'Sullivan et al., 2011 (6.48) Metereau & Dreher, 2013 (5.80) Gershman et al., 2009 (2.57) Murray et al., 2008 (0.98)

The studies contributing to each cluster and the extent of their contribution (as a percentage) to the overall cluster are marked. SF = superficial subregion of amygdala

reinforcement and also included a high contribution from TD studies. There were relatively few TD studies, but these employed either monetary or liquid reinforcers, and about half were Pavlovian designs. In general, given the small number of such studies (Pavlovian/TD/liquid) and the potential for confounding, the findings from these maps should be interpreted cautiously.

Both the individual-related striatal and the fixed-related midbrain activations were predominantly collected from instrumental rather than Pavlovian studies, as would be expected from the higher proportion of instrumental studies. The striatal activations associated with individual studies were elicited half by monetary and half by other reinforcers, whereas the midbrain activation associated with fixed studies was also represented by studies employing a variety of different reinforcers.

Pavlovian versus instrumental (Table 5)

The instrumental RPE map was similar to the overall RPE map, aside from the lack of midbrain activation. Striatal activations were slightly more medial than the overall RPE cluster and did not extend as convincingly into the lateral striatum (putamen), nor farther into the amygdala. In addition, the left caudate was activated in this contrast. By contrast, the Pavlovian studies yielded two clusters in the left putamen/amygdala and right amygdala. The amygdala activations were predominantly located in the superficial subregion.

Bilateral amygdala and left lateral putamen were significantly more likely to be activated in Pavlovian than in instrumental paradigms. The reverse contrast yielded a significant cluster in the left caudate (anterior and dorsally located), as well as smaller activations in more ventral regions of the medial striatum. A small region reflecting the conjunction of instrumental and Pavlovian tasks was apparent in the left putamen.

Fixed versus individual (Table 6)

The individual map was also similar to the overall RPE map, without the presence of the midbrain cluster or any activation within the dorsal striatum. The striatal activations were focused within the medial regions of the ventral striatum. By contrast, the fixed map yielded two clusters: one in left putamen and one in the midbrain. Statistical comparison of the contrasts yielded greater activation in the bilateral ventral striatum (medially focused) for the individual contrast, as well as the left operculum and left visual cortex. The fixed contrast yielded a large midbrain cluster, as well as very small differences in the left lateral putamen. A cluster representing the conjunction of fixed and individual was present in the left putamen.

PE at outcome versus TD error (Table 7)

Studies that modeled PE only at the US made up a large proportion of the data, and consequently the US PE map was very similar to the overall RPE map. The seven TD error studies yielded a cluster including the left lateral striatum (putamen) and amygdala. A conjunction between the two was again observed within the left putamen. The TD error studies showed activated left amygdala/hippocampus more than did the US PE studies, whereas the latter showed greater activation in the left caudate and left frontal operculum.

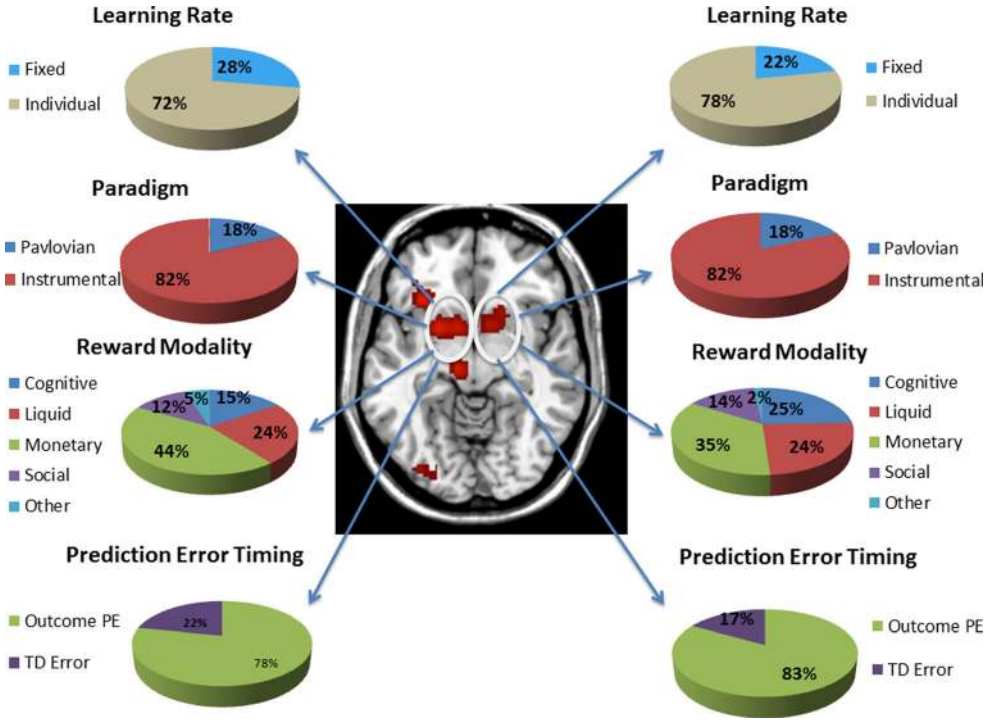

Fig. 3 Map of significant ALE clusters associated with the reward prediction error contrast, with activations in the striatum circled. Pie charts show the contributions of the studies of a particular class to the bilateral striatum activation. Percentages are not corrected for base rate

Reinforcer type (Table 8)

As with the outcome PE map, monetary reinforcement occurred frequently in the selection of studies. Thus, the monetary subanalysis revealed a pattern of activations very similar to the overall RPE contrast. The other reinforcer-type subanalyses were somewhat underpowered, and we did not perform statistical contrasts of these maps. The cognitive subanalysis did not reveal any

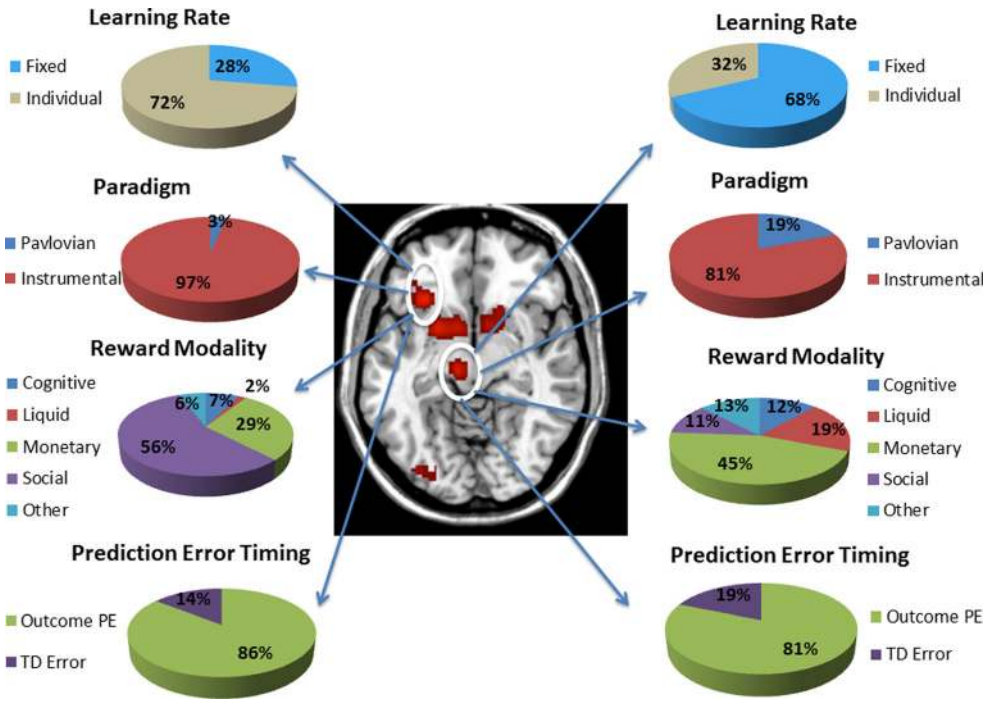

Fig. 4 Map of significant ALE clusters associated with the reward prediction error contrast, with activations in the midbrain and frontal operculum circled. Pie charts show the contributions of the studies of a particular class to each activation. Percentages are not corrected for base rate

Table 5 ALE clusters representing instrumental (Instr) and Pavlovian (Pav) activations, including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinates	Size
Instrumental			
Left putamen	5.96	–16 6 –12	597
Left ventral caudate	5.46	–10 8 –6	
Left dorsal caudate (head)	4.64	–12 8 8	
Right ventral striatum	4.78	14 6 –14	397
	4.52	18 16 –6	
	3.37	6 18 –4	
Left frontal operculum	6.32	–32 24 –8	233
Left fusiform gyrus (V4), inferior occipital, lingual gyrus	4.21	–22 –82 –18	162
	4.17	–34 –84 –8	
	3.93	–24 –88 –16	
Pavlovian			
Left putamen/amygdala (SF)	5.18	–24 4 –10	194
	4.06	–20 0 –22	
Right amygdala (SF)	5.16	26 –2 –12	136
	3.71	36 0 –10	
	3.66	38 –2 –8	
Pav/Instr Conjunction: Left putamen	4.77	–22 6 –12	50
Instr > Pav: Left caudate	2.98	–10 8 10	58
	2.95	–8 4 10	
	2.34	–10 4 16	
Instr > Pav: Left pallidum	1.93	–12 4 –2	29
	1.86	–8 2 –4	
	1.74	–6 4 –2	
Pav > Instr: Right amygdala (SF/LB)	2.89	24 –8 –8	112
	2.49	34 –2 –12	
Pav > Instr: Left putamen, left amygdala (SF)	2.69	–28 2 –10	82
Pav > Instr: Left amygdala (SF/LB), left hippocampus (EC)	2.35	–22 2 –20	50

SF = superficial subregion of amygdala; LB = laterobasal subregion of amygdala; EC = entorhinal cortex

Table 6 ALE clusters representing individual (Ind) and fixed activations, including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinates	Size
Individual			
Left ventral striatum	6.13	–18 4 –12	441
	5.14	–10 10 –6	
Right ventral striatum	4.78	18 8 –4	415
	4.64	14 6 –16	
	4.25	10 8 –10	
	3.88	24 0 –12	
	3.54	6 18 –4	
Left fusiform gyrus (V4), inferior occipital, lingual gyrus	4.38	–34 –84 –8	217
	4.06	–24 –88 –16	
	3.96	–24 –84 –18	
	3.72	–26 –88 –8	
	3.47	–24 –82 –8	
Left frontal operculum	6.20	–30 24 –8	166
Fixed			
Midbrain/thalamus	5.44	–8 –22 –6	278
	3.65	6 –16 –10	
Left putamen (lateral)	4.57	–24 6 –8	111
Fixed/Ind Conjunction: Left putamen	4.21	–24 6 –10	51
Ind > Fixed: Left inferior occipital, fusiform gyrus (V4)	2.80	–34 –80 –8	119
	2.77	–36 –80 –12	
	2.60	–24 –80 –6	
	2.23	–28 –88 –8	
Ind > Fixed: Left ventral striatum	2.44	–12 6 –10	113
	2.35	–10 10 12	
Ind > Fixed: Right ventral striatum	2.50	20 8 –8	53
Ind > Fixed: Left frontal operculum	2.09	–26 28 –4	40
	2.00	–28 24 –6	
Fixed > Ind: Midbrain/thalamus	2.62	–4 –24 –4	151
	2.47	–2 –12 –10	
	2.46	–10 –26 –6	

significant clusters, but the liquid and social reinforcement maps yielded several distinct clusters. Liquid rewards elicited lateral putamen and amygdala activations, whereas social rewards produced two left hemispheric activations: One was similar to the frontal opercular/insula cluster in the main reward PE contrast; the second was in the left inferior parietal cortex.

High versus low smoothing (Table 9)

High-smoothing studies were associated with bilateral putamen and amygdala activation, as well as activation in the left frontal operculum. Low-smoothing studies were associated with the thalamus/midbrain and left frontal operculum. The opercular activations were not similar enough to yield a significant conjunction. High-smoothing studies were significantly more likely to activate the right amygdala than were low-smoothing studies. The low-smoothing studies were more likely to activate a small cluster of the thalamus, toward the top of the midbrain/thalamus cluster identified in the main RPE contrast.

Overall conjunction

A conjunction analysis was conducted across all of the main contrast types (Pavlovian/instrumental, fixed/individual, RW/TD, high/low smoothing) using the minimum statistic across the cluster-thresholded contrasts for each of the eight maps (Rottschy et al., 2012). A 30-voxel cluster was revealed in the left putamen (–22, 6, 9) across the first three pairs of contrasts (i.e., excluding smoothing). This cluster thus reflects the strongest convergent evidence for a neural correlate of a signed RPE signal that we were able to obtain (see Fig. 5). However, when the smoothing-related contrasts were included, no clusters were identified.

Expected value (Table 10)

The ALE analysis of studies reporting EV yielded a single activation in the subgenual anterior cingulate cortex (ACC; Table 10, Fig. 6). To illustrate specificity, the RPE and EV maps were contrasted. The subgenual ACC was significantly more likely to be activated in the EV than in the RPE

Table 7 ALE clusters representing temporal difference (TD) error and prediction error (PE) at outcome activations, including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinate	Size
TD error			
Left putamen, amygdala (SF/LB), hippocampus	5.12	–16 6 –14	270
	4.31	–24 6 –10	
	4.20	–20 0 –22	
	3.69	–28 –8 30	
PE at outcome			
Left ventral striatum	5.45	–10 8 –6	566
	5.21	–20 6 –12	
	4.62	–12 8 8	
Right ventral striatum	4.59	18 8 –4	365
	4.52	18 16 –6	
	4.35	10 8 –10	
	3.44	6 18 –4	
Midbrain/thalamus	5.10	–8 –20 –6	115
Left frontal operculum	6.28	–32 24 –8	240
PE at outcome only/TD error conjunction: Left putamen	4.74	–18 6 –12	112
	4.31	–24 6 –10	
TD error > Outcome PE: Left Amygdala (SF, LB), hippocampus (EC)	3.95	–18 2 –24	127
	3.26	–18 0 –28	
	2.64	–16 –6 –30	
Outcome PE > TD error: Left caudate	3.30	–10 10 6	126
	2.97	–8 4 10	
	2.95	–10 8 10	
	2.51	–10 8 14	
	1.93	–8 10 0	
Outcome PE > TD error: Left frontal operculum, inferior frontal gyrus, pars orbitalis	2.47	–40 34 –10	64
	2.01	–34 32 –12	
	1.98	–34 32 –8	
	1.97	–36 36 –12	
	1.77	–38 26 –12	

SF = superficial subregion of amygdala; LB = laterobasal subregion of amygdala; EC = entorhinal cortex

condition, whereas the left striatum and midbrain were significantly more likely to be activated in the RPE than in the EV condition. No significant clusters representing the conjunction of EV and RPE were observed.

Discussion

In line with previous animal and human studies, the present meta-analysis confirmed our core hypotheses: that the midbrain and striatum represented reward prediction errors, whereas the subgenual cingulate—a caudal region of the vmPFC—represents expected value. In addition, this meta-analysis revealed that the frontal operculum and visual cortices are part of the RPE network, mainly recruited during social rewards and attentional processing, respectively. Although these results are largely compatible with previous meta-analyses of the neural bases of PEs (Garrison et al., 2013), reward anticipation and receipt (Diekhof, Kaps,

Table 8 ALE clusters representing the activations associated with different reinforcers, including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinate	Size
Monetary			
Left ventral striatum	6.07	–18 6 –14	278
Left inferior occipital, lingual gyrus (V4)	4.87	–34 –84 –8	215
	4.24	–24 –86 –16	
	3.25	–26 –98 –12	
Right ventral striatum	4.35	10 10 –10	278
	3.99	16 6 –14	
	3.311	18 16 –6	
Liquid			
Left putamen/amygdala (SF, LB)	5.76	–24 4 –10	260
	4.37	–28 –2 –14	
Right amygdala (SF, LB, CM)	5.30	26 –2 –12	154
	3.71	38 –2 –8	
	3.43	32 –14 –14	
Social			
Left frontal operculum/IFG	5.74	–30 24 –10	234
Left inferior parietal lobule (hIP1, inferior parietal cortex (PGa, PFm)	4.25	–40 –54 42	123
	3.92	–50 –56 42	
Cognitive			
No regions			

SF = superficial subregion of amygdala; LB = laterobasal subregion of amygdala; CM = centromedial subregion of amygdala; EC = entorhinal cortex

Table 9 ALE clusters representing activations associated with high and low smoothing kernels, including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinate	Size
High Smoothing			
Left putamen, amygdala	6.40	–20 6 –12	524
	3.61	–28 –4 –16	
Right putamen, amygdala	4.78	26 –2 –12	430
	4.66	14 6 –14	
	4.11	20 10 –4	
	3.55	34 2 –12	
	3.11	6 4 4	
Left frontal operculum	5.55	–30 24 –8	137
Low Smoothing			
Thalamus/midbrain	4.81	–8 –18 –2	112
Left inferior frontal gyrus (pars orbitalis), frontal operculum	4.13	–34 28 –12	109
	4.07	–36 22 –6	
	3.85	–30 28 –14	
High/Low Smoothing Conjunction	–	–	–
High > Low: Right amygdala (SF)	2.44	24 –2 –14	57
	1.99	14 0 –16	
	1.97	16 2 –14	
Low > High: Left thalamus	2.09	–6 –18 –2	46

SF = superficial subregion of amygdala.

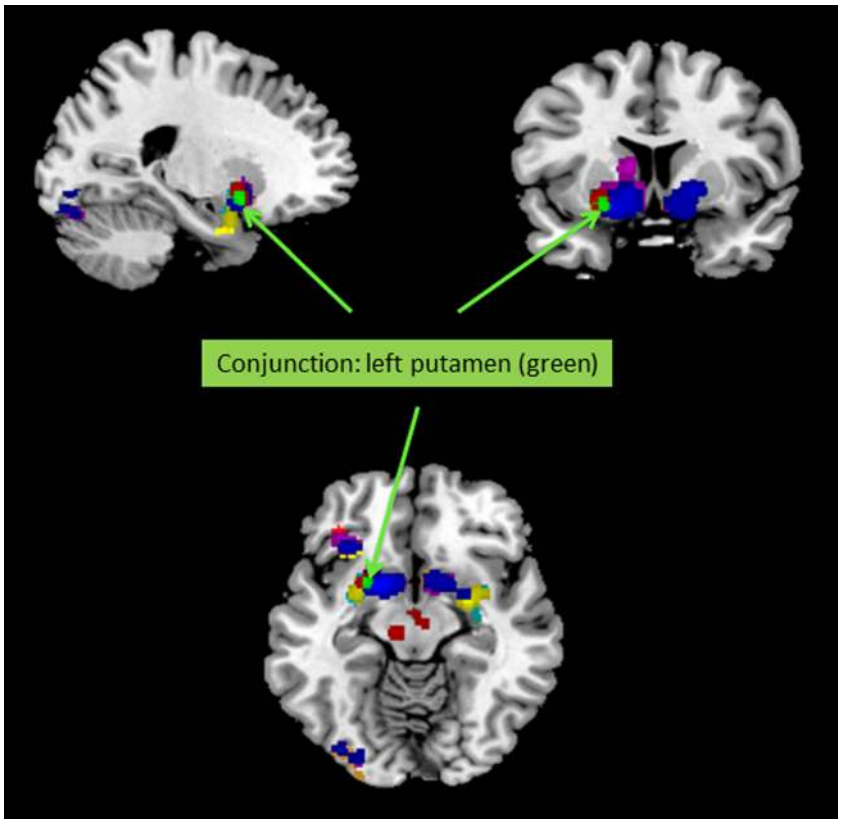

Fig. 5 Conjunction map showing overlap of the ALE maps from individual subgroup analyses (fixed, individual, Pavlovian, instrumental, outcome PE, TD, monetary, liquid, and social), with the left putamen cluster ($x = -22$, $y = 6$, $z = 9$, cluster size = 30) from the conjunction analysis marked with arrows

Falkai, & Gruber, 2012; Liu, Hairston, Schrier, & Fan, 2011; Sescousse et al., 2013), and value (Bartra et al., 2013; Clithero & Rangel, 2014; Levy & Glimcher, 2012; Peters & Buchel, 2010), the present study extends this work by focusing exclusively on the neural correlates of parametric RPEs and EV derived from reinforcement learning models. We identified methodological factors that might have contributed to the divergent findings, including Pavlovian/instrumental designs, reinforcer type, and smoothing kernel size.

Core PE network

The reproducibility of fMRI BOLD images is often a concern, with test–retest reliability of the method being generally modest, and very poor in some cases (Bennett & Miller, 2010). Moreover, methodological differences across studies, including differences between scanners, paradigms, participants, and analysis software may further conspire to amplify between-study heterogeneity. Nevertheless, a core network of regions associated with PEs was readily identified, including the ventral striatum and midbrain, as predicted. Indeed, even for two regions that were not predicted—the left frontal operculum and left visual cortex—over ten studies contributed to each of these clusters. This suggests that this core PE network is robust to between-study variability and reflects a level of specificity of the activations. However, each of the activations should be interpreted carefully; it is often difficult to distinguish certain psychological events, due to a shared but spurious correlation with the general linear model regressor. The variability of paradigms may act to provide some

Table 10 ALE cluster representing the activation associated with expected value (EV), including peak *t* statistics, MNI coordinates, and cluster sizes

Region	*t* Statistic	Coordinate	Size	Studies Participating (Percentage Contribution)
Subgenual cingulate	4.85 3.54	4 34 –6 –6 28 –20	172	FitzGerald et al., 2012 (26.52) Wunderlich et al., 2010 (24.44) Glascher et al., 2009 (21.24) Bernacer et al., 2013 (13.99) Kim et al., 2006 (9.83) Klein et al., 2007 (2.80) Takemura et al., 2011 (0.69)

The studies contribution to the cluster, and their percentage contributions, are marked

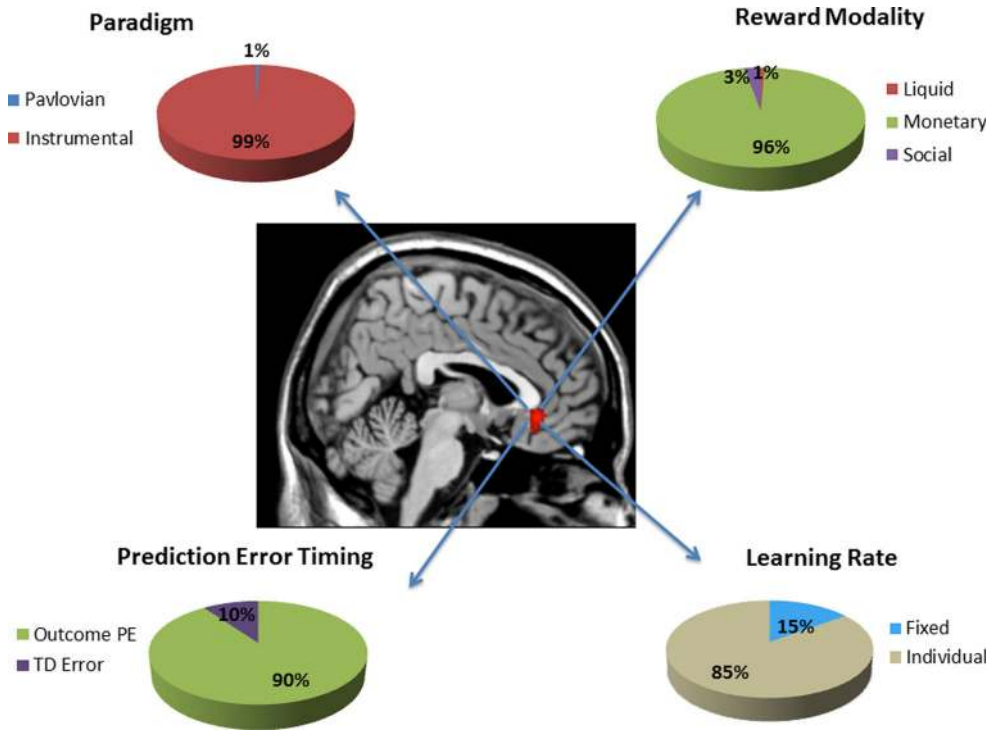

Fig. 6 Map of significant ALE clusters associated with the estimated value contrast. Pie charts show the contributions of the studies of a particular class to the subgenual cingulate activation. Percentages are not corrected for base rate

decorrelation of irrelevant variables from the RPE construct. For example, the lack of PE signals in the medial PFC is consistent with animal electrophysiological studies (Roesch et al., 2010), although medial OFC activation has been shown to be coupled to RPE in some human fMRI studies. Our findings are consistent with the view that this is likely to be due to the correlation inherent between appetitive properties of the outcome and RPE in many of these designs (Erdeniz, Rohe, Done, & Seidler, 2013; Rohe et al., 2012).

Aside from the reinforcement learning signal hypothetically encoded by dopamine-rich regions such as the midbrain and ventral striatum, associative learning algorithms are often extended to account for salience and attentional phenomena. These constructs may be necessary for interpreting RPE correlates in the visual cortex, amygdala, and insula. For example, the Pearce–Hall (PH) model (Pearce & Hall, 1980) emphasizes that the cues associated with surprising outcomes command attention: PEs not only strengthen associations, but a similar signal, reflecting surprise associated with the outcome, may control the rate at which such associations are strengthened. In the PH model, stimuli that are accompanied by larger PEs attract attention, and thus become more readily associated with other stimuli. A recent theme has been to argue that a PH signal might be coupled to the surprising outcome itself, rather than to conditioned stimuli. For example, a recent study by Li, Schiller, Schoenbaum, Phelps, and Daw (2011) suggested that, consistent with animal learning studies (Maddux, Kerfoot, Chatterjee, & Holland, 2007), the amygdala codes surprise, as predicted by the PH model, rather than a signed RPE signal.

In the present study, we found amygdala activation coupled to the RPE contrast. In the probabilistic designs that are widely used, it would be difficult to dissociate a PH signal from the basic RPE contrast. It may then be that RPE-coupled amygdala activation reflects some confounding of a PH signal with the RPE signal, particularly because a PH parameter is often not concurrently modeled. However, amygdala activation was particularly associated with studies in which liquid was used as a reinforcer, whereas larger smoothing kernels were also associated with greater activation in the amygdala. These factors should be independent of the learning rule and contingency under investigation, and should be adequately controlled in future studies of the PH rule.

Other regions that have played a well-established role in attention in the fMRI literature were also coupled to the RPE contrast, including the left visual cortex. Although reward-related responses in the visual cortex have been identified, a recent study argued that these signals may reflect attentional processing rather than the appetitive and dopamine-related properties of the reward (Arsenault, Nelissen, Jarraya, & Vanduffel, 2013). With the RPE contrast, we also identified a left frontal operculum/anterior insula region that is activated by a wide range of stimuli and task designs, and thus perhaps has a general role in task set representation (Dosenbach et al.,

2006). Nevertheless, the activation of this region by reward has been quite well characterized. A study by Rutledge, Dean, Caplin, and Glimcher (2010) parametrically manipulated the reward probabilities of wins and losses, finding that the response of the anterior insula to reward did not follow a pattern that would be expected from a PE signal. It was, however, modulated to some degree by the probability of the outcome, insofar as activation was not observed in the region if the outcome was fully predicted, and showed fairly consistent activation across wins and losses if the outcome was uncertain. Given that the paradigms in the present study have generally included a degree of outcome uncertainty, this opens the possibility that anterior insula activation may become coupled with an RPE regressor, while not accurately reflecting the predicted RPE signal. Less obvious is the fact that paradigms employing social reinforcement were particularly able to elicit activation in this region. An interpretation of the Rutledge et al. study might suggest that this is simply related to the kind of contingencies employed in the social paradigms, but equally it is worth considering the possibility that the anterior insula may play a distinct role in the reinforcement process itself.

Pavlovian versus instrumental

Although the majority of studies have been instrumental, requiring participants to make a choice, we contrasted these studies with a small number of Pavlovian designs. We found differential activation in the left caudate (dorsal striatum), consistent with an influential study by O'Doherty and colleagues (2004) in which the striatum was argued to follow the "actor–critic" model: the anterior, dorsal caudate ("actor") was engaged when behavior output was required. By contrast, the ventral striatum ("critic") was engaged during errors of value prediction, whether or not a response was required to obtain reward. This distinction is also broadly consistent with animal lesion studies, since the dorsomedial striatum of rodents—a likely homologue of the caudate region identified in the present study and that of O'Doherty et al. (2004)—plays a key role in instrumental, goal-directed behavior (Yin, Ostlund, Knowlton, & Balleine, 2005), whereas the ventral striatum is more consistently implicated in Pavlovian behaviors (Corbit & Balleine, 2011; Parkinson, Olmstead, Burns, Robbins, & Everitt, 1999).

Although the notion that the striatum contributes to action selection in a manner predicted by the actor–critic model has steadily gathered currency, it was somewhat undermined by a previous meta-analysis by Garrison and colleagues (2013). This study showed that, although both the dorsal and ventral striatum were engaged by instrumental designs, both were significantly more activated by these designs than by Pavlovian designs. Our findings contrast with that study, since we did find significant activation in the ventral striatum elicited by Pavlovian designs, although it was somewhat more lateral than the equivalent activations seen in instrumental designs.

Together, the present study and that of Garrison et al. (2013) may provoke further debate about the success of the actor–critic model as an account of the striatum's influence on behavior. However, there are several important reasons why providing a definitive contribution to this question might be difficult. First, it has been noted (e.g., Coricelli et al., 2005; Yeung, Holroyd, & Cohen, 2005) that designs in which a (human) participant is required to make a choice, and is reinforced for doing so, are potentially more engaging than Pavlovian designs, and consequently can provide more robust neural signals. Given that the magnetic resonance scanner requires that an individual lie for long periods in a darkened room, performing an often repetitive task, this consideration is not to be taken lightly, and can make it difficult to design an effective Pavlovian paradigm. This may explain both the preponderance of instrumental tasks in the literature and the second key limitation—that Pavlovian designs tend to focus on liquid reinforcers rather than other domains. This is presumably because liquid is a powerful primary reinforcer, particularly when the participant is thirsty (e.g., Kumar et al., 2008), and this may somewhat compensate for the potential lack of engagement described above. A final limitation is the nature of the definition of instrumental and Pavlovian designs. Instrumental behavior can be defined on the basis of the contingency between a particular action and an outcome (Balleine & Dickinson, 1998), and the manner in which a participant can use this information to obtain reinforcement. The presence of stimuli in all of the paradigms that we considered in the present work complicates this issue somewhat. Specifically, in any of the instrumental designs included in the present work, it cannot be assumed that this action–outcome contingency was the sole factor that determined choice. Rather, an individual's responses may also have been susceptible to influence by the presented stimuli and by the relationships between the stimuli and reinforcement.

Fixed versus individual learning rates

We investigated whether the strategy of reinforcement learning model fitting, upon which the pattern of the RPE (and EV) regressors was based, was associated with different patterns of neural activation. Although across most situations the patterns of RPEs associated with fixed and individual model fitting should be highly similar, it is nevertheless unclear exactly how sensitive the pattern of activations is to the parameterization of the underlying model. Daw (2011) has consistently argued that the fixed (or, more particularly, group fixed) strategy offers advantages over estimating the model parameters per individual. On the other hand, regarding the fitting of models to behavioral data, Estes and Maddox (2005)

have argued that individual-participant fitting avoids certain sources of bias associated with group averaging.

The fixed subgroup showed the strongest corroboration of the classic RPE hypothesis pioneered by Schultz and colleagues (Schultz et al., 1997), since the midbrain was engaged in these studies. In addition, activation in the lateral putamen was also observed, as would be expected on the basis of anatomical connectivity (Haber et al., 2000). However, if the individual method was suboptimal, we would not expect the method to have obtained traction in the literature—individual studies being more common than fixed ones—and more importantly, we would not expect a distinct pattern of activations to emerge. It is possible to imagine various scenarios in which the presence of suboptimal acquisition or preprocessing parameters that impair the detection of midbrain activations would sustain the observation of a certain pattern of weaker ventral striatal RPE-associated responses beyond the canonical network, but even then, the focus of the activation should not show such a reproducibly medial focus within the striatum. It also does not seem likely that a suboptimal RPE regressor would be better coupled to an experimental confound, such as the response to the reward itself (Rohe et al., 2012). Within the reinforcement learning framework we have set out, the most likely remaining explanation is that the neural responses to RPEs generated by different learning rates are reflected across different regions of the brain (Glascher & Buchel, 2005). For example, a model by M. J. Frank, Moustafa, Haughey, Curran, and Hutchison (2007) distinguished a rapid but time-dependent learning mechanism, ascribed to the OFC, and a slower, incremental learning mechanism, ascribed to the striatum. Both mechanisms used similar RW-based learning rules, although more recent, comparable models have employed a working-memory-based system rather than a rapid reinforcement learning system (Collins & Frank, 2012). This might, therefore, provide one interpretation of our data, with the modification that the medial striatum encodes a more variable learning rate (across individuals), perhaps better linked to trial-by-trial choice performance, whereas the midbrain and lateral putamen reflect a more homogeneous, slower learning rate that is not as strongly reflected in behavior.

Conjunction analyses

A further level of specificity is afforded by the conjunction analysis examining which regions have been identified across different designs, and thus are relatively invariant. Across several of the subgroup analyses (i.e., fixed/individual, Pavlovian/instrumental, and RW/TD), the left putamen was identified. The region was notable insofar as it was positioned at the midpoint between the classic ventromedial striatal region, which may correspond to the nucleus accumbens in humans (Haber & Knutson, 2010), and a more clearly lateralized putamen region. Given that these two regions may be anatomically distinct (Haber et al., 2000), it is important to consider the extent to which smoothing may have played a part in this finding. The smoothing of individual participant images is considered to be an important preprocessing step: Though not without drawbacks, the method is thought to enhance statistical power, by increasing the ratio of signal to noise (Yue, Loh, & Lindquist, 2010), and increases the underlying smoothness for Gaussian random field-based (cluster) analyses (Hayasaka & Nichols, 2003). It is intriguing that one subgrouping analysis that did not yield activation in this region was the conjunction of studies that used high and low smoothing kernels. In a recent study, Sacchet and Knutson (2013) demonstrated that larger smoothing kernels can influence the localization of peak activation within the ventral striatum, with larger kernels yielding more posterior activations. In our study, the variability in the magnitudes of smoothing kernels across studies was relatively small, with the large majority of studies choosing an 8-mm kernel, and no significant differences between the low/high smoothing subgroups were seen. However, it was also notable that studies using a small smoothing kernel were (nonsignificantly) more capable of revealing midbrain activation. Given that the midbrain is a small structure, matched filter theory (for fMRI, see Yue et al., 2010) would predict that a smaller filter should therefore be advantageous to identify activation in this region. Overall, as was suggested by Sacchet and Knutson, differences in smoothing across studies may provide significant additional heterogeneity, and alternative smoothing methods that honor the geometry and sizes of these regions may be valuable in future studies.

Core expected value network

Our meta-analysis of reinforcement learning studies of EV identified a subregion of the subgenual cingulate cortex, corresponding most closely to areas 25 and 32 of the human and monkey vmPFC. This phylogenetically ancient agranular region is likely homologous to the paralimbic and infralimbic cortex of rodents (Wallis, 2012).

At the first approximation, our findings converge with primate electrophysiological (Kennerley et al., 2009; Kennerley & Wallis, 2009a, 2009b; Morrison & Salzman, 2009; Padoa-Schioppa & Assad, 2006, 2008; Roesch & Olson, 2004, 2005; Wallis & Miller, 2003) and lesion (Izquierdo, Suda, & Murray, 2004; Noonan et al., 2010; Rudebeck & Murray, 2011) studies, as well as rodent lesion studies (Gallagher, McMahan, & Schoenbaum, 1999; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011; Takahashi et al., 2009), implicating the OFC in value computations. Yet, the substantial anatomical heterogeneity between these literatures cannot be ignored. Most primate electrophysiological studies have recorded value signals from more rostral, central orbitofrontal regions

(BAs 11 and 13). Rodent studies have often employed lesions of the more rostral and lateral OFC (Gallagher et al., 1999; McDannald et al., 2011; Takahashi et al., 2009). In contrast, our subgenual cingulate cluster is more medial and caudal and does not extend to the orbital surface. This discrepancy was recently discussed by Wallis (2012), who pointed out a few possible solutions to this puzzle. First, rostromedial OFC BOLD activations in BA 11, medial BA 13, and ventral BA 10 are obscured by the susceptibility artifact. Thus, value signals in the human brain may well extend into the rostral and central OFC areas highlighted by primate physiological studies. However, a recent meta-analysis of fMRI studies of reward value that was not limited to reinforcement learning studies, by Bartra and colleagues (2013), reported value-related activations in the medial rostral OFC areas most affected by the susceptibility artifact, but not in the more lateral central OFC, in which signal is often better preserved.

Another set of considerations stems from the medial–lateral organization of the orbitofrontal circuits (Ongur & Price, 2000). The lateral, "orbital" circuit of Carmichael and Price (1996) encompasses central OFC areas, which integrate sensory inputs carrying information about extrinsic food values: taste, olfaction, and vision. It is often argued that this lateral circuit represents not only the values of foods and liquids typically used in animal experiments, but those of external stimuli and outcomes in general (Schoenbaum, Takahashi, Liu, & McDannald, 2011; Wallis, 2012). Physiologists have typically recorded from this circuit in their studies of primate and rodent OFC (Kennerley et al., 2009; Kennerley & Wallis, 2009a, 2009b; Morrison & Salzman, 2009; Padoa-Schioppa & Assad, 2006, 2008; Roesch & Olson, 2004, 2005; Wallis & Miller, 2003).

An additional reason why fMRI studies may have not detected value signals in central OFC is its diametrically opposed value-encoding scheme (Wallis, 2012): Some OFC neurons increase and others decrease their firing rates in response to increasing value (Kennerley & Wallis, 2009a; Morrison & Salzman, 2009; Padoa-Schioppa & Assad, 2006). These opposing responses may cancel each other out at the level of the BOLD signal. The medial orbital circuit, encompassing the vmPFC and the subgenual cingulate in particular, has prominent visceral and motor connections (Carmichael & Price, 1996; Ongur & Price, 2000). Its putative functions include sensing internal states, tracking social value, and bridging outcome value and action selection (Bouret & Richmond, 2010; Noonan et al., 2010; Rudebeck et al., 2008; Rudebeck, Buckley, Walton, & Rushworth, 2006). Grabenhorst and Rolls (2011) have placed the vmPFC downstream from the OFC in the processing of reward signals, proposing that the vmPFC receives stimulus value information from the OFC, incorporates other variables such as cost into the decision, and transmits it to motor areas. VmPFC responses often scale with subjective pleasure, which may best correspond to the reward rate or the total value of the contingencies that can be exploited.

Not only are the findings of vmPFC value signals consistent in human fMRI studies, but they are also less well established in the primate electrophysiological literature (Wallis, 2012; but see Strait, Blanchard, & Hayden, 2014). This discrepancy may reflect methodological differences between the human and monkey studies. For example, human studies have mostly used secondary reinforcers such as money and correct/incorrect feedback. Only 2/16 value studies in our meta-analysis used primary rewards (liquid). One of them detected value signals in the vmPFC (Takemura, Samejima, Vogels, Sakagami, & Okuda, 2011), and one did not (Gradin et al., 2011), and neither found value signals in the central OFC. Furthermore, the meta-analysis by Bartra and colleagues (2013) reported vmPFC value signals for both primary and monetary rewards. A similar explanation focuses on the putative predilection of the vmPFC for social value signals (Rudebeck et al., 2006). The presence of vmPFC value signals in fMRI studies that have used primary, nonsocial rewards argues against this explanation. That said, demand characteristics may be a confound in human imaging studies of value signals, and experimenters may thus need to conceal contingency manipulations. In summary, our finding of reinforcement-learning-estimated value signals in the vmPFC/subgenual cingulate is consistent with non-reinforcement-learning-based human imaging studies and diverges somewhat from the primate electrophysiological studies, which have tended to find value signals in the central OFC.

Given that the EV map was restricted to the vmPFC, a supplementary conjunction analysis of the RPE and EV contrasts did not reveal significant results. Given that the EV maps reflect future expected rewards, it is plausible that a TD-related signal should be observed at this stage, and thus a concurrent striatal or midbrain activation. In fact, significantly different activations were observed between the RPE network (RPE > EV) and the vmPFC EV cluster (EV > RPE). A statistical account of this observation may relate to the combined inclusion of RPE and EV regressors in the general linear model used in the analysis of many of the studies: The presence of each regressor concurrently, combined with a suitable design, may act to orthogonalize these two events and distinguish the resulting maps. Nevertheless, our findings are also consistent with the view that a phasic TD signal might be distinct (in this case, neuroanatomically) from an EV signal (Ludvig, Sutton, & Kehoe, 2008).

Limitations

Although striking consistency in the patterns of activation was observed across paradigms, there was nevertheless evidence of different classes of paradigms leading to different patterns of findings, as we discussed. A limitation of the inferences that can be drawn from analyses of these differences was caused by the presence of confounds between different categories. This was particularly acute for Pavlovian–TD–liquid designs, because of their relative infrequency. In particular, amygdala RPE-coupled

activations were associated with these classes of designs, making it difficult to draw strong conclusions about the amygdala's engagement by a paradigm class. Overall, our method of contrasting paradigm classes required that all other dimensions be controlled for strong inferences to be obtained. Although this was not possible, the findings nevertheless point to particular trends of experimental design that may precipitate differences in the patterns of neural activation obtained.

Refutations or refinements of reinforcement learning models are of course a crucial part of their theoretical development within neuroscientific investigation (Gamez, 2012). However, we have restricted our analysis to studies in which the reinforcement learning model was not refuted or otherwise argued to be an inferior account of the pattern of data, albeit we did allow for some modifications of parameterization to the basic RW or TD model. Bayesian models such as the Bayesian learner (Behrens, Woolrich, Walton, & Rushworth, 2007), hidden-Markov models (Hampton, Bossaerts, & O'Doherty, 2006), and Bayesian reinforcement learning (Mathys, Daunizeau, Friston, & Stephan, 2011), as well as the Kalman filter (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006), can all exhibit advantages over many of the models we have examined in the present work. However, the superior performance of the alternative models in the studies that we opted to exclude may have been a result of peculiarities of the experimental designs, which might render these studies more heterogeneous a priori, and thus less suitable for meta-analysis. In addition, the nature of this advantage should be carefully qualified (Myung, 2000): Often, these models are representationally more powerful, perhaps reflecting inherent features of the experimental design (e.g., the rule transitions embedded within reversal learning: Behrens et al., 2007; Hampton et al., 2006). Although pursuing the benefits of these models is likely to be a topic of major ongoing interest, we argue that the incremental increase in complexity and representational capacity of many of these models creates a natural, qualitative distinction from the more traditional reinforcement learning methods that provided the focus of the present work.

Another limitation of the present study involves the limitation of meta-analysis, over and above the direct pooling of data within a "mega"-analysis. A judicious combination of fMRI studies of conditioning could in theory be performed, perhaps along similar lines to the analysis of task-related neural activation by Dosenbach and colleagues (2006). If possible, this would certainly afford a more direct contrast of different modeling strategies (e.g., fixed/individual learning rates, smoothing kernels), and possibly also of procedural differences (e.g., reinforcer types, response contingencies). Moreover, this approach may afford more detailed investigation of the relationships between individual functional activations and anatomy, providing that adequate structural data are available. The overlap between individually defined regions of interest and brain activations would diminish the necessity of spatial smoothing and potentially increase the specificity in regions of high between-participant anatomical variation.

We also restricted our study inclusion to healthy adult groups. Individual differences in a variety of demographic factors can influence the patterns of reinforcement-learning-related neural activation and represent possible unmeasured sources of intersubject variability. Again, a "mega"-analysis with suitably recorded data might provide some control of these effects. However, the consistency of some of our findings (e.g., left putamen) across methodological dimensions suggests that these factors may serve to modulate a core pattern of activation rather than to yield qualitative differences. Overall, because ALE has been argued to be statistically conservative (Graham et al., 2013), it is likely that our findings broadly represent a central, reproducible motif that may provide a useful reference point for future studies of reinforcement learning and reward-based conditioning studies. Indeed, an increase in the number of available reinforcement learning studies would allow greater power to address the full diversity of reinforcement-learning-related processes in the human brain. Although the number of studies available was adequate, further information could be usefully gleaned by increasing the number of studies (e.g., Rottschy et al., 2012), particularly if they provided data from designs not well represented in the present selection (e.g., liquid–TD studies).

Summary

In the present work, we have identified a pattern of human neural correlates of RPE and EV signals derived from simple reinforcement learning algorithms. Our findings accord well with the existing literature, particularly with electrophysiological studies of experimental animals, in our identification of dopamine-rich regions such as the midbrain and striatum in RPE signaling, and the ventromedial prefrontal cortex in EV representation. The main contribution of the present work has been to demonstrate that various methodological factors can influence the patterns of findings. These include factors that are possible to control at the analysis stage (e.g., learning rate estimation, smoothing), but also factors that must be examined experimentally (e.g., reinforcer type, behavioral output). Overall, the reinforcement learning framework has been an empirically successful paradigm for investigating the neurobiology of appetitive behavior, and we anticipate that a new generation of studies will seek to develop the implications of these findings further.

Author note The authors declare no financial conflicts of interest that may have biased the present work. We thank the following individuals, and their associated research teams, for contributing the data used in the study: Rumana Chowdhury, Jessica Cohen, Thomas Fitzgerald, Gerhard Jocham, Rebecca Jones, Andreas Heinz, Thorsten Kahnt, John O'Doherty, Soyoung Park, Oliver Robinson, Florian Schlagenhauf, and Wouter van den Bos. We also thank Masahiko Haruno, Noreen O'Sullivan, Ben Seymour, and Craig Stark for answering our questions.

References

Arsenault, J. T., Nelissen, K., Jarraya, B., & Vanduffel, W. (2013). Dopaminergic reward signals selectively decrease fMRI activity in primate visual cortex. *Neuron, 77,* 1174–1186. doi:10.1016/j.neuron.2013.01.008

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37,* 407–419.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76,* 412–427. doi:10.1016/j.neuroimage.2013.02.063

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10,* 1214–1221. doi:10.1038/nn1954

Bellebaum, C., Jokisch, D., Gizewski, E. R., Forsting, M., & Daum, I. (2012). The neural coding of expected and unexpected monetary performance outcomes: Dissociations between active and observational learning. *Behavioural Brain Research, 227,* 241–251. doi:10.1016/j.bbr.2011.10.042

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences, 1191,* 133–155. doi:10.1111/j.1749-6632.2010.05446.x

Bernacer, J., Corlett, P. R., Ramachandra, P., McFarlane, B., Turner, D. C., Clark, L., & Murray, G. K. (2013). Methamphetamine-induced disruption of frontostriatal reward learning signals: Relation to psychotic symptoms. *American Journal of Psychiatry, 170,* 1326–1334. doi:10.1176/appi.ajp.2013.12070978

Bouret, S., & Richmond, B. J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *Journal of Neuroscience, 30,* 8591–8601. doi:10.1523/JNEUROSCI. 0049-10.2010

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology, 97,* 3036–3045. doi:10.1152/jn.01211.2006

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. *Cerebral Cortex, 18,* 1485–1495. doi:10.1093/cercor/bhm198

Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review, 58,* 413–423. doi:10.1037/H0054576

Bush, R. R., & Mosteller, F. (1953). A stochastic model with applications to learning. *Annals of Mathematical Statistics, 24,* 559–585. doi:10.1214/aoms/1177728914

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology, 371,* 179–207. doi:10.1002/(SICI)1096-9861(19960722)371:2<179::AID-CNE1>3.0.CO;2-#

Chiu, P. H., Lohrenz, T. M., & Montague, P. R. (2008). Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nature Neuroscience, 11,* 514–520. doi:10.1038/nn2067

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Duzel, E., & Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nature Neuroscience, 16,* 648–653. doi:10.1038/nn.3364

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience, 9,* 1289–1302. doi:10.1093/scan/nst106

Cohen, M. X. (2007). Individual differences and the neural representations of reward expectation and reward prediction error. *Social Cognitive and Affective Neuroscience, 2,* 20–30. doi:10.1093/scan/nsl021

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35,* 1024–1035. doi:10.1111/j.1460-9568.2011.07980.x

Corbit, L. H., & Balleine, B. W. (2011). The general and outcome-specific forms of Pavlovian–instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *Journal of Neuroscience, 31,* 11786–11794. doi:10.1523/JNEUROSCI. 2711-11.2011

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8,* 1255–1262. doi:10.1038/nn1514

Critchley, H. D., & Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *Journal of Neurophysiology, 75,* 1673–1686.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science, 319,* 1264–1267.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making, affect, and learning: Attention and performance XXIII* (pp. 3–38). Oxford, UK: Oxford University Press.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876–879. doi:10.1038/nature04766

Dayan, P., & Walton, M. E. (2012). A step-by-step guide to dopamine. *Biological Psychiatry, 71,* 842–843. doi:10.1016/j.biopsych.2012.03.008

Diekhof, E. K., Kaps, L., Falkai, P., & Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude—An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia, 50,* 1252–1266. doi:10.1016/j.neuropsychologia.2012.02.007

Dombrovski, A. Y., Szanto, K., Clark, L., Reynolds, C. F., III, & Siegle, G. J. (2013). Reward signals, attempted suicide, and impulsivity in late-life depression. *JAMA Psychiatry, 70,* 1020–1030. doi:10.1001/jamapsychiatry.2013.75

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., & Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron, 50,* 799–812. doi:10.1016/j.neuron.2006.04.031

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage, 59,* 2349–2361. doi:10.1016/j.neuroimage.2011.09.017

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,* 938–949. doi:10.1016/j.neuroimage.2011.05.021

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,* 2907–2926. doi:10.1002/hbm.20718

Erdeniz, B., Rohe, T., Done, J., & Seidler, R. D. (2013). A simple solution for model comparison in bold imaging: The special case of reward prediction error and reward outcomes. *Frontiers in Neuroscience, 7,* 116. doi:10.3389/fnins.2013.00116

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12,* 403–408.

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. *Frontiers in Neuroscience, 6,* 148. doi:10.3389/fnins.2012.00148

Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: Evidence from humans with frontal lobe damage. *Annals of the New York Academy of Sciences, 1239,* 51–58. doi:10.1111/j.1749-6632.2011.06229.x

FitzGerald, T. H., Friston, K. J., & Dolan, R. J. (2012). Action-specific value signals in reward-related regions of the human brain. *Journal of Neuroscience, 32,* 16417–16423. doi:10.1523/JNEUROSCI.3254-12.2012

Frank, G. K., Reynolds, J. R., Shott, M. E., & O'Reilly, R. C. (2011). Altered temporal difference learning in bulimia nervosa. *Biological Psychiatry, 70,* 728–735. doi:10.1016/j.biopsych.2011.05.011

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience, 17,* 51–72. doi:10.1162/0898929052880093

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, 104,* 16311–16316.

Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *Journal of Neuroscience, 19,* 6610–6614.

Gamez, D. (2012). From Baconian to Popperian neuroscience. *Neural Systems and Circuits, 2,* 2. doi:10.1186/2042-1001-2-2

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 37,* 1297–1310. doi:10.1016/j.neubiorev.2013.03.023

Gershman, S. J., Pesaran, B., & Daw, N. D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *Journal of Neuroscience, 29,* 13524–13531. doi:10.1523/JNEUROSCI.2469-09.2009

Glascher, J., & Buchel, C. (2005). Formal learning theory dissociates brain regions with different temporal integration. *Neuron, 47,* 295–306. doi:10.1016/j.neuron.2005.06.008

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

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences, 15,* 56–67. doi:10.1016/j.tics.2010.12.004

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., & Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain, 134,* 1751–1764. doi:10.1093/brain/awr059

Graham, J., Salimi-Khorshidi, G., Hagan, C., Walsh, N., Goodyer, I., Lennox, B., & Suckling, J. (2013). Meta-analytic evidence for neuroimaging models of depression: State or trait? *Journal of Affective Disorders, 151,* 423–431. doi:10.1016/j.jad.2013.07.002

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience, 20,* 2369–2382.

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology, 35,* 4–26. doi:10.1038/npp.2009.129

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26,* 8360–8367. doi:10.1523/JNEUROSCI.1010-06.2006

Hayasaka, S., & Nichols, T. E. (2003). Validating cluster size inference: Random field and permutation methods. *NeuroImage, 20,* 2343–2356.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences, 13,* 517–523. doi:10.1016/j.tics.2009.09.004

Holroyd, C. B., & Coles, M. G. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex, 44,* 548–559. doi:10.1016/j.cortex.2007.08.013

Howard-Jones, P. A., Bogacz, R., Yoo, J. H., Leonards, U., & Demetriou, S. (2010). The neural mechanisms of learning from competitors. *NeuroImage, 53,* 790–799. doi:10.1016/j.neuroimage.2010.06.027

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience, 24,* 7540–7548. doi:10.1523/JNEUROSCI.1921-04.2004

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,* 1606–1613. doi:10.1523/JNEUROSCI.3904-10.2011

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., & Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *Journal of Neuroscience, 31,* 13039–13045. doi:10.1523/JNEUROSCI.2972-11.2011

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,* 1332–1345. doi:10.1162/jocn.2009.21092

Kamin, L. J. (1968). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York, NY: Appleton-Century-Crofts.

Kennerley, S. W., Dahmubed, A. F., Lara, A. H., & Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *Journal of Cognitive Neuroscience, 21,* 1162–1178. doi:10.1162/jocn.2009.21100

Kennerley, S. W., & Wallis, J. D. (2009a). Encoding of reward and space during a working memory task in the orbitofrontal cortex and anterior cingulate sulcus. *Journal of Neurophysiology, 102,* 3352–3364. doi:10.1152/jn.00273.2009

Kennerley, S. W., & Wallis, J. D. (2009b). Evaluating choices by single neurons in the frontal lobe: Outcome value encoded across multiple decision variables. *European Journal of Neuroscience, 29,* 2061–2073. doi:10.1111/j.1460-9568.2009.06743.x

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,* e233. doi:10.1371/journal.pbio.0040233

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., & Ullsperger, M. (2007). Genetically determined differences in learning from errors. *Science, 318,* 1642–1645. doi:10.1126/science.1145044

Kobayashi, S., Pinto de Carvalho, O., & Schultz, W. (2010). Adaptation of reward sensitivity in orbitofrontal neurons. *Journal of Neuroscience, 30,* 534–544. doi:10.1523/JNEUROSCI.4009-09.2010

Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How we learn to make decisions: Rapid propagation of reinforcement learning prediction errors in humans. *Journal of Cognitive Neuroscience, 26,* 635–644. doi:10.1162/jocn_a_00509

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,* 2084–2093.

Lea, S. (1978). The psychology and economics of demand. *Psychological Bulletin, 85,* 441–466. doi:10.1037/0033-2909.85.3.441

Leathers, M. L., & Olson, C. R. (2012). In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science, 338,* 132–135. doi:10.1126/science.1226405

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology, 22,* 1027–1038. doi:10.1016/j.conb.2012.06.001

Li, J., McClure, S. M., King-Casas, B., & Montague, P. R. (2006). Policy adjustment in a dynamic economic game. *PLoS ONE, 1,* e103. doi:10.1371/journal.pone.0000103

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience, 14,* 1250–1252. doi:10.1038/nn.2904

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 35,* 1219–1236. doi:10.1016/j.neubiorev.2010.12.012

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging, 22,* 1517–1531. doi:10.1016/j.mri.2004.10.018

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2008). Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Computation, 20,* 3034–3054. doi:10.1162/neco.2008.11-07-654

Maddux, J. M., Kerfoot, E. C., Chatterjee, S., & Holland, P. C. (2007). Dissociation of attention in learning and action: Effects of lesions of the amygdala central nucleus, medial prefrontal cortex, and posterior parietal cortex. *Behavioral Neuroscience, 121,* 63–79. doi:10.1037/0735-7044.121.1.63

Madlon-Kay, S., Pesaran, B., & Daw, N. D. (2013). Action selection in multi-effector decision making. *NeuroImage, 70,* 66–79. doi:10.1016/j.neuroimage.2012.12.001

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience, 5,* 39. doi:10.3389/fnhum.2011.00039

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *Journal of Neuroscience, 31,* 2700–2705. doi:10.1523/jneurosci.5499-10.2011

Metereau, E., & Dreher, J. C. (2013). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex, 23,* 477–487. doi:10.1093/cercor/bhs037

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla–Wagner model. *Psychological Bulletin, 117,* 363–386.

Morrison, S. E., & Salzman, C. D. (2009). The convergence of information about rewarding and aversive stimuli in single neurons. *Journal of Neuroscience, 29,* 11471–11483. doi:10.1523/Jneurosci.1815-09.2009

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., & Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Molecular Psychiatry, 13*(239), 267–276. doi:10.1038/sj.mp.4002058

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204. doi:10.1006/jmps.1999.1283

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience, 32,* 551–562. doi:10.1523/JNEUROSCI.5498-10.2012

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., & Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences, 107,* 20547–20552. doi:10.1073/pnas.1012246107

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304,* 452–454. doi:10.1126/science.1094285

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,* 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,* 2825–2835. doi:10.1016/j.neuropsychologia.2011.06.008

Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10,* 206–219.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441,* 223–226.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience, 11,* 95–102. doi:10.1038/nn2020

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,* 7749–7753. doi:10.1523/JNEUROSCI.5587-09.2010

Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., & Everitt, B. J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *Journal of Neuroscience, 19,* 2401–2411.

Paxinos, G., & Huang, X.-F. (1995). *Atlas of the human brain stem*. San Diego, CA: Academic Press.

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. doi:10.1037/0033-295X.87.6.532

Peters, J., & Buchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research, 213,* 135–141. doi:10.1016/j.bbr.2010.04.031

Petrides, M., & Pandya, D. (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 17–58). Amsterdam, The Netherlands: Elsevier.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400,* 233–238. doi:10.1038/22268

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Robinson, O. J., Overstreet, C., Charney, D. R., Vytal, K., & Grillon, C. (2013). Stress increases aversive prediction error signal in the ventral striatum. *Proceedings of the National Academy of Sciences, 110,* 4129–4133. doi:10.1073/pnas.1213923110

Rodriguez, P. F. (2009). Stimulus-outcome learnability differentially activates anterior cingulate and hippocampus at feedback processing. *Learning and Memory, 16,* 324–331. doi:10.1101/lm.1191609

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,* 306–313. doi:10.1002/hbm.20186

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. doi:10.1152/jn.00173.2010

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience, 10,* 1615–1624. doi:10.1038/nn2013

Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science, 304,* 307–310. doi:10.1126/science.1093223

Roesch, M. R., & Olson, C. R. (2005). Neuronal activity in primate orbitofrontal cortex reflects the value of time. *Journal of Neurophysiology, 94,* 2457–2471. doi:10.1152/jn.00373.2005

Rohe, T., Weber, B., & Fliessbach, K. (2012). Dissociation of BOLD responses to reward prediction errors and reward receipt by a model comparison. *European Journal of Neuroscience, 36,* 2376–2382. doi:10.1111/j.1460-9568.2012.08125.x

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage, 60,* 830–846. doi:10.1016/j.neuroimage.2011.11.050

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. S. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *Journal of Neuroscience, 28,* 13775–13785. doi:10.1523/jneurosci. 3541-08.2008

Rudebeck, P. H., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science, 313,* 1310–1312. doi:10.1126/science.1128197

Rudebeck, P. H., & Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *Journal of Neuroscience, 31,* 10569–10578. doi:10.1523/Jneurosci. 0091-11.2011

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,* 13525–13536. doi:10.1523/JNEUROSCI. 1747-10.2010

Sacchet, M. D., & Knutson, B. (2013). Spatial smoothing systematically biases the localization of reward-related brain activity. *NeuroImage, 66,* 270–277. doi:10.1016/j.neuroimage.2012.10.056

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science, 310,* 1337–1340.

Schlagenhauf, F., Rapp, M. A., Huys, Q. J., Beck, A., Wustenberg, T., Deserno, L., & Heinz, A. (2012). Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Human Brain Mapping*. doi:10.1002/hbm.22000

Schoenbaum, G., Takahashi, Y., Liu, T. L., & McDannald, M. A. (2011). Does the orbitofrontal cortex signal value? *Critical Contributions of the Orbitofrontal Cortex to Behavior, 1239,* 87–99. doi:10.1111/j.1749-6632.2011.06210.x

Schonberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27,* 12860–12867. doi:10.1523/JNEUROSCI. 2496-07.2007

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,* 772–781. doi:10.1016/j.neuroimage.2009.08.011

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599. doi:10.1126/science.275.5306.1593

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,* 644–656. doi:10.1016/j.neuroimage.2009.11.083

Sescousse, G., Caldu, X., Segura, B., & Dreher, J. C. (2013). Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 37,* 681–696. doi:10.1016/j.neubiorev.2013.02.002

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,* 1234–1240. doi:10.1038/nn1527

Simmons, J. M., Ravel, S., Shidara, M., & Richmond, B. J. (2007). A comparison of reward-contingent neuronal activity in monkey orbitofrontal cortex and ventral striatum: Guiding actions toward rewards. *Annals of the New York Academy of Sciences, 1121,* 376–394. doi:10.1196/annals.1401.028

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron, 82,* 1357–1366. doi:10.1016/j.neuron.2014.04.032

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, UK: Cambridge Univ Press.

Takahashi, Y. K., Roesch, M. R., Stalnaker, T. A., Haney, R. Z., Calu, D. J., Taylor, A. R., & Schoenbaum, G. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron, 62,* 269–280. doi:10.1016/j.neuron.2009.03.005

Takemura, H., Samejima, K., Vogels, R., Sakagami, M., & Okuda, J. (2011). Stimulus-dependent adjustment of reward prediction error in the midbrain. *PLoS One, 6,* e28337. doi:10.1371/journal.pone.0028337

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,* 1233–1241. doi:10.1016/j.neunet.2006.05.039

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *Journal of Neuroscience, 23,* 10402–10410.

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,* 301–310. doi:10.1152/jn.00762.2005

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature, 398,* 704–708. doi:10.1038/19525

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,* 765–780.

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping, 33,* 1–13. doi:10.1002/hbm.21186

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,* 3384–3391. doi:10.1152/jn.91195.2008

van den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral Cortex, 22,* 1247–1255. doi:10.1093/cercor/bhr198

Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences, 27,* 468–474. doi:10.1016/j.tins.2004.06.006

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412,* 43–48.

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience, 15,* 13–19. doi:10.1038/nn.2956

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience, 18,* 2069–2081. doi:10.1046/j.1460-9568.2003.02922.x

Watanabe, N., Sakagami, M., & Haruno, M. (2013). Reward prediction error signal enhanced by striatum-amygdala interaction explains the acceleration of probabilistic reward learning by emotion. *Journal of Neuroscience, 33,* 4487–4493. doi:10.1523/JNEUROSCI.3400-12.2013

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron, 58,* 967–973. doi:10.1016/j.neuron.2008.04.027

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2010). Economic choices can be made using only stimulus values. *Proceedings of the National Academy of Sciences, 107,* 15005–15010. doi:10.1073/pnas.1002258107

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex, 15,* 535–544.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22,* 513–523. doi:10.1111/j.1460-9568.2005.04218.x

Yue, Y., Loh, J. M., & Lindquist, M. A. (2010). Adaptive spatial smoothing of fMRI images. *Statistics and its Interface, 3,* 3–13.