NeuroImage 89 (2014) 1–9

Contents lists available at ScienceDirect

NeuroImage

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

Fatal attraction: Ventral striatum predicts costly choice errors in humans

J.R. Chumbley *, P.N. Tobler, E. Fehr

Laboratory for Social and Neural Systems, University of Zurich, Switzerland

ARTICLE INFO

Article history:
Accepted 18 November 2013
Available online 26 November 2013

ABSTRACT

Animals approach rewards and cues associated with reward, even when this behavior is irrelevant or detrimental to the attainment of these rewards. Motivated by these findings we study the biology of financially-costly approach behavior in humans. Our subjects passively learned to predict the occurrence of erotic rewards. We show that neuronal responses in ventral striatum during this Pavlovian learning task stably predict an individual's general tendency towards financially-costly approach behavior in an active choice task several months later. Our data suggest that approach behavior may prevent some individuals from acting in their own interests.

Introduction

Optimal goal-directed choice involves learning which option is best and picking it. Problems may arise from having learned an inaccurate 'model' of the action-outcome contingencies or from failing to consult any model at all, as when previously-reinforced habits thoughtlessly persist in new settings (Balleine and O'Doherty, 2009). Hypothetically, choice errors might also arise from a behavioral attraction to options with positive associations, even when this attraction is costly and has never been reinforced or habitualized. It is notable here that non-human animals approach reward-predictive cues even when this approach is inefficient or specifically non-reinforced,[1] thereby providing the canonical evidence for hardwired "Pavlovian" approach systems (Domjan et al., 2000; Locurto, 1981; Rosenthal and Matthews, 1978; Sanabria et al., 2006; Williams and Williams, 1969). However, it is currently unclear whether this tendency can produce financially-costly choice errors in humans and, if so, by what mechanism. Specifically, it is unclear if humans will select the wrong cue in a financial choice task because of behaviorally-irrelevant reward associations previously acquired in a separate and strictly passive Pavlovian learning setting.

It is widely believed that passive reward learning depends on dopaminergic reward prediction-errors (RPEs) in the ventral striatum (VS) (O'Doherty et al., 2006; Schultz et al., 1997). Critically however, this region is also implicated in innate Pavlovian behaviors (Balleine and Killcross, 1994; Berridge, 2007; Berridge and Robinson, 1998; Boureau and Dayan, 2010; Corbit and Balleine, 2003; Hall et al., 2001; Ikemoto and Panksepp, 1999; Killcross et al., 1997; Mogenson et al., 1980; Panksepp, 2004; Parkinson et al., 2000; Reynolds and Berridge, 2001; Salamone and Correa, 2002; Sesack and Grace, 2009; Talmi et al., 2008) but *not* goal-directed (Balleine and Killcross, 1994; Corbit et al., 2001) or habitual behaviors (Reading et al., 1991; Robbins et al., 1990). This makes the VS relevant to the acquisition of inflexible behavioral attraction towards cues with positive associations.

In our task, male subjects passively observed associations between two neutral cues and un/rewarding pictures, respectively: pictures of fractals versus pictures of bikini-clad women posing in provocative body postures. Months later, the same subjects had to choose between the very same two predictive cues for money, in the absence of any fractal or erotic outcomes. They could earn more money by choosing the correct cue or choose the erotically-associated cue for considerably less money. This task therefore pits Pavlovian value against instrumental value — the correct cue lacks erotic associations but has a higher expected financial return. It therefore resembles both 'Pavlovian–instrumental-transfer' (PIT) paradigms and 'negative auto-maintenance' paradigms (Balleine, 2005; Guitart-Masip et al., 2010; Prévost et al., 2012; Sanabria et al., 2006; Talmi et al., 2008). Using fMRI, we aimed to predict individual vulnerability to 'fatal' Pavlovian attraction from passive VS learning responses. We used responses in the ventral striatum as a proxy for Pavlovian learning, as has been well-established in other studies (cf. (O'Doherty et al., 2004; Tobler et al., 2007) and further references therein. Our behavioral control studies imply the effectiveness of this learning (see below).

Materials and methods

Experiment I

Participants

All 17 male subjects (20–25 year, median age 22) had normal or corrected-to-normal vision and were screened to exclude those with a previous history of neurological or psychiatric disease. All gave informed consent and the study was approved by the Research Ethics Committee of the Canton of Zurich.

Study overview

During the passive fMRI learning task, subjects viewed cue — (erotic) reward associations on a visual display. We call this experiment the *fMRI*

* Corresponding author at: Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland. Fax: +41 44 634 49 07.
E-mail address: justin.chumbley@econ.uzh.ch (J.R. Chumbley).

[1] Aka (negative) auto-maintenance, sign-tracking, and autoshaping.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.11.039

session because we recorded the subjects' brain activation during this task with functional magnetic resonance imaging (fMRI).

After the fMRI session the same subjects attended a *behavioral session*, in a different lab and location. This latter consisted of two *phases* containing new cues and old fMRI cues respectively: together these phases allowed us to assess the stability of any Pavlovian–instrumental effects over different cues and times.

fMRI task

After completing a consent form and MR safety questionnaire, participants were invited to read the task instructions. Subjects then passively viewed the cue–reward associations in Fig. 1 on a computer screen. Each cue was mostly followed by its respective outcome: an erotic versus fractal image. (On such trials, this image was selected randomly from a corpus of 10 erotic/10 fractal images, with replacement.) On a minority 25% of trials this outcome was omitted (see Fig. 1). In total, subjects were presented each cue 50 times in a random order. As a convention, we will refer to the cue with erotic associations as cue A, and the cue with factual associations as cue B.

Each passive trial started with a variable ITI with only a fixation cross visible in the center of the screen. Each trial commenced randomly between 4 and 6 s after the end of the previous trial. Specifically, the ITI was drawn from a uniform distribution on the 2 second interval beginning at 4 s after the end of the previous trial. The ITI was followed by the presentation of one out of two visual cues, at random, for 3 s. The outcome – erotic or fractal image – was superimposed on this cue for the final 1.5 s of presentation (a forward or 'delay' conditioning procedure). All rewards were independent samples from the conditional distributions shown in Fig. 1. In each trial the cue was drawn randomly with probability 1/2. Cues were counterbalanced between subjects.

Behavioral task

After a 2–4 month interval subjects visited a different behavioral laboratory ~1 km from the location of the scanner. There were two phases to the behavioral task in which subjects either passively acquired a new association (phase 1) or re-acquired the old association (phase 2). In each phase they also made instrumental choices from money, as elaborated below. The written instructions for this behavioral task are given in the Supplementary Fig. 4.

Old cues (phase 2). Subjects alternated between passive associative learning blocks and active choice-learning blocks with the cues previously encountered in fMRI (5, 10-trial blocks of each). The first block was always passive. These passive or 'rest' blocks respected the same contingencies as the fMRI study but here cue A and cue B were presented simultaneously, on the LEFT and the RIGHT of subjects' computer screen, randomized trial-by-trial (Fig. 2a). Erotic versus fractal images were then superimposed on cue A and cue B, respectively. These passive trials served to renew and sustain previously-established associations. Active blocks required subjects to choose between ambiguous lotteries — i.e. lotteries for which subjects had no a priori information about their probability of success (Ellsberg, 1961). They therefore had to use trial-and-error to learn the best option, i.e. option B (Fig. 2b and below). On each trial, each lottery yielded zero or one point (1 point = 0.2 CHF). Written instructions invited subjects to earn as many points as possible during active blocks and did not indicate that erotic or fractal images would be presented: No erotic or fractal images were presented during active blocks. After selecting one option, subjects were informed whether or not they had successfully collected a point: "1" or "0" displayed respectively.

Critically, in each active choice trial the two options were indicated by cue A and cue B (Fig. 2b). In other words, the subjects had to choose between the same 2 cues presented on passive trials. If they chose cue A (previously associated with an erotic picture) they earned 1 point with probability 0.3, while if they chose cue B (previously associated with an unrewarding fractal) they earned 1 point with probability 0.7. Thus cue B pits Pavlovian value against instrumental value: it lacks erotic associations but has a higher expected return in the choice task.

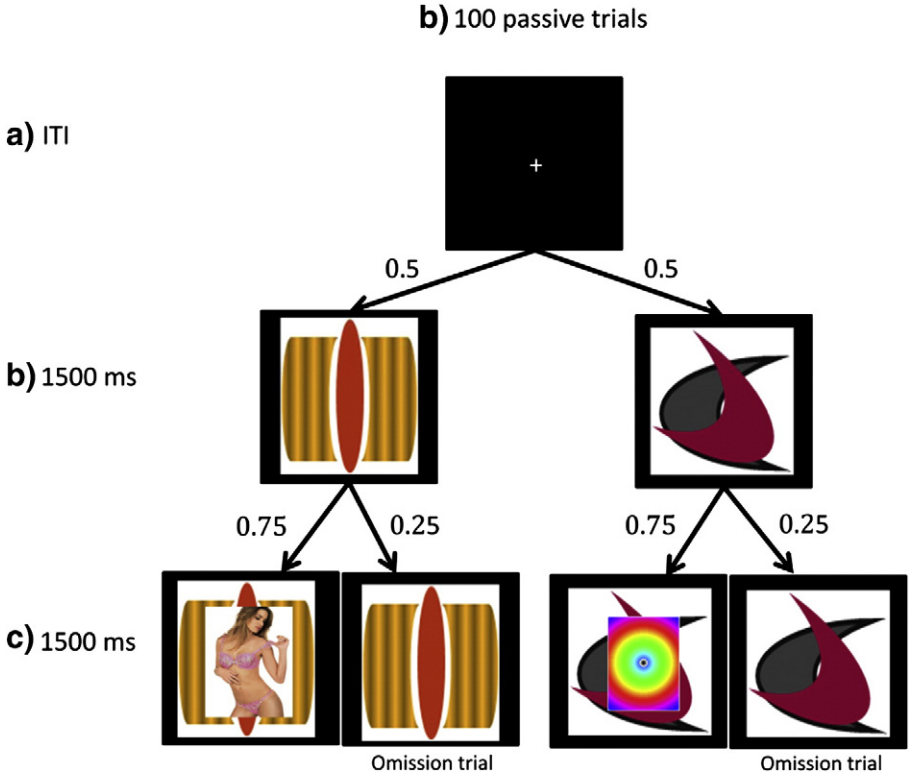

Fig. 1. fMRI associative learning. Subjects passively observed stimulus–reward pairings on the computer screen. Cue A (left) predicted erotic images, cue B predicted fractals, both with 0.75 probability. Cues were counterbalanced over subjects. For a specific and statistically powerful index of RPE learning signal, we contrasted fMRI responses to the omission of erotic versus fractal images (see text).

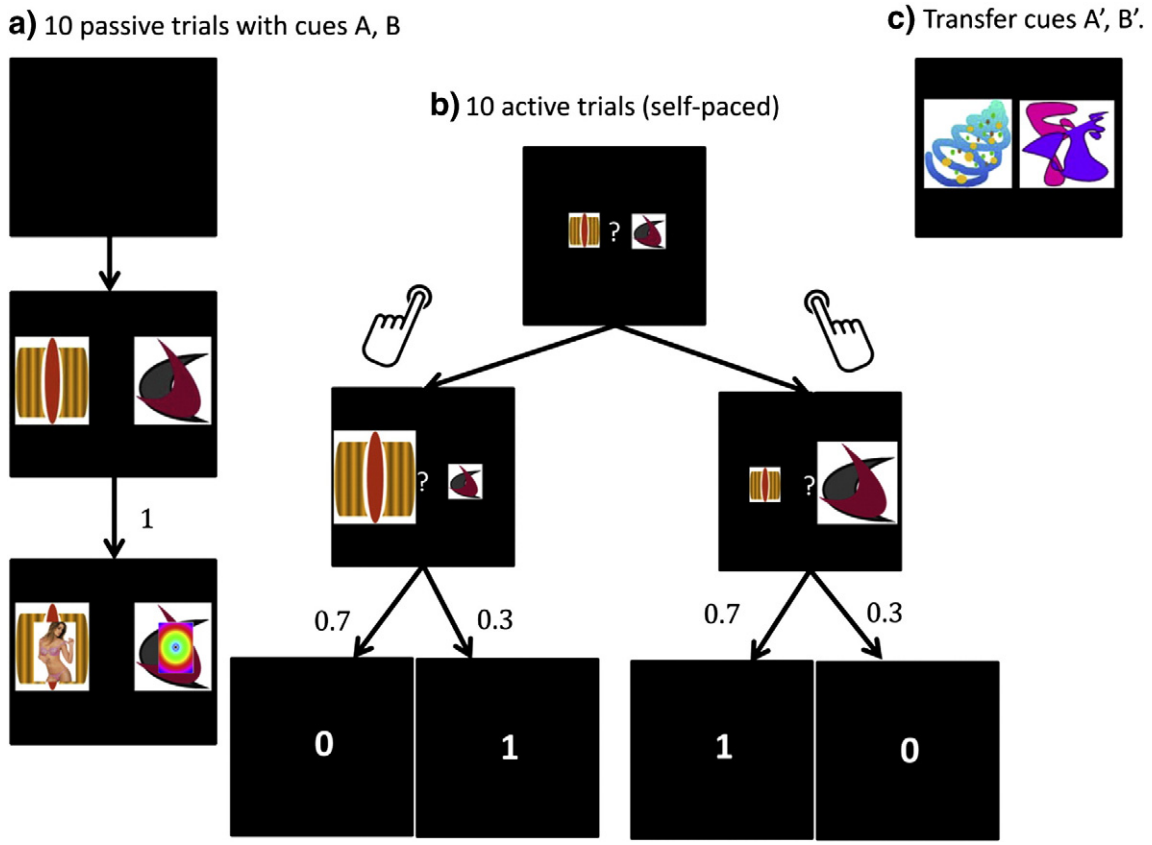

Fig. 2. Follow-up behavioral task. a. Passive blocks had the same associative contingencies as in Fig. 1 but cues A and B were presented simultaneously. b. Passive blocks were followed by active choice blocks, arranged such that the attractive cue A had lower expected payoff, i.e. 30% probability rather than 70% of earning 1 point (worth roughly USD 0.2). c. In the first session of the behavioral follow up, these 'transfer' cues A′ and B′ play the role of cues A and B described above. Contingencies in this session were otherwise the same as described in panels (a) and (b).

Selecting either option by button-press instantly magnified the corresponding cue on the subjects' screen (Fig. 2b). In this way we operationalized financially-costly 'approach behavior' towards positive cues. If subjects aim to maximize their financial payoff, then choosing cue B is optimal and choosing cue A is suboptimal. It is in this specific sense that we use the term "choice error" throughout this work.

New cues (phase 1). Before completing the above behavioral task with cues A and B, subjects first completed a phase with new cues A′ and B′, not present in the fMRI. This phase also comprised 5 passive blocks alternated with 5 active blocks and had a structure exactly analogous to that described above. Cue A′ had the same contingencies as cue A: it was a low-value choice with erotic associations during passive associative training. Similarly cues B and B′ had identical contingencies. In this phase subjects could acquire fresh associations to new cues. We call these new cues "transfer cues" (Fig. 2b) because they enable us to examine whether the ventral striatal RPE when learning with cues A and B can predict subjects' choice errors to different cues A′ and B′. Thus, a subject prone to financially-costly errors under the original cues A and B should acquire similar impairment to new cues A′ and B′. This would indicate the stability of a learning mechanism.

Imaging parameters

Images were acquired using a Philips Achieva 3T whole-body scanner with an 8 channel SENSE head coil (Philips Medical Systems, Best, The Netherlands) at the Laboratory for Social and Neural Systems Research (SNS Lab), Zurich. Subjects viewed the stimuli through a mirror fitted on top of the head coil. We acquired gradient echo T2*-weighted echo-planar images (EPIs) with blood–oxygen-level-dependent (BOLD) contrast (slices/volume, 37; repetition time, 2.47 s). Approximately 350 volumes were collected in each session of the experiment. Scan onset times varied randomly relative to stimulus onset times. Volumes were acquired at a +15° tilt to the anterior commissure–posterior commissure line, rostral > caudal. Imaging parameters were the following: echo time, 30 ms; field of view, 220 mm. The spatial resolution of our functional data is 3 × 3 × 3 mm. A T1-weighted 3D-TFE high-resolution structural image was also acquired for each participant. For this, the following parameters were used: TR = 7.4 s, TE = 3.4 s, TI = 876.2 ms (minimum TI delay), flip angle (deg) = 8, FOV = 250 × 250 (×180), matrix size = 240 (reconstruction matrix), voxel size = 1 × 1 × 1 (1.041 reconstructed); and acquisition time 5.57 min.

fMRI preprocessing

Statistical parametric mapping (SPM8; Functional Imaging Laboratory, University College London) served to spatially realign functional data, coregister them to the individual anatomical image before normalizing to standard MNI space and smoothing with an isometric Gaussian kernel with a full-width at a half-maximum of 9 mm.

fMRI statistical analysis

Overview. We hypothesized that a stronger tendency to (re)acquire positive associations in some subjects would lead them to approach and select the low-value option (i.e. cue A) more often. In addition, we were interested in whether subjects' vulnerability to (re)acquiring this 'approach bias' is predicted by the ventral striatum learning signals measured in the fMRI session several months earlier. This would suggest that the potency of their historical learning, as indicated by the size of their RPE, predicted the (re)acquisition of the approach behavior at follow-up. We did two analyses: one relatively model-independent, and one inspired by a classical learning model.

Model-independent analysis

We asked if the magnitude of individuals' negative RPE in the ventral striatum predicted how often they choose the low-return option with erotic associations in the behavioral follow-up. For each subject, we used linear regression to quantify this (negative) 'RPE': the relative fMRI BOLD decrements to omitted erotic images versus omitted fractals. We choose this definition of RPE because these two trial-types are equally surprising and have identical sensory features between subjects: in contrast non-omission trials have outcomes with different sensory features i.e. women versus fractals. In addition this contrast on negative RPEs has greater statistical sensitivity for identifying prediction errors (see Discussion). This contrast, hereafter RPE, discloses a specific response to the omission of anticipated rewards at each voxel in the brain.

First-level design (within-subject). We used a standard rapid-event-related fMRI approach in which evoked hemodynamic responses y to each stimulus event (cue, outcome, omission) are estimated separately convolving a canonical hemodynamic response function with a stimulus function encoding the onsets for each event. We entered these into a design matrix along with 6 variables m coding potential movement confounds. The corresponding general linear model has the form

$$y = X\theta + \varepsilon$$

with $X = [\text{cue A}, w_+, w_-, \text{cue B}, f_+, f_-, m, 1]$. Here w_+ codes the presentation of an erotic female image: w_- codes the omission of this image (similarly for *cue B* and fractal images f). The error ε is modeled in the standard way as an autoregressive 1 process. In Fig. 3 we consider the contrast of omitted erotic versus fractal images

$$v = c^T\left(X^TX\right)^{-1}Xy$$

where $c = [0\,0\,1\,0\,0\,-\,1\,0\,0\ldots]$.

Second-level design (between-subject). We used the standard summary-statistic approach for inference. Namely, we treated subject i's first-level contrast image v_i as an observation. To examine its relation to subjects' approach behavior we used a simple linear regression.

$$v = Z\phi + \varepsilon$$

where $Z = [1, b]$, b is the total number of financially-costly approaches and the *i.i.d.* ε term now describes unexplained between-subject variability.

Model-dependent analysis

In a separate and complementary model-based analysis, we attempted to predict subjects' behavioral error rate from the scale of prediction errors defined formally (Rescorla and Wagner, 1972). In particular we initialized to 0 two Rescorla–Wagner (RW) learning models corresponding to the cue A—reward association and cue B—fractal association. These learning models have the form $\eta_{t+1} = \eta_t + \alpha\delta$, where $\delta = (x - \eta_t)$ is the prediction error and prescribes how each cue's predictive association with x is updated from trial t to trial $t + 1$. While RW was initially conceived as a model for learning the predictive *value* of a

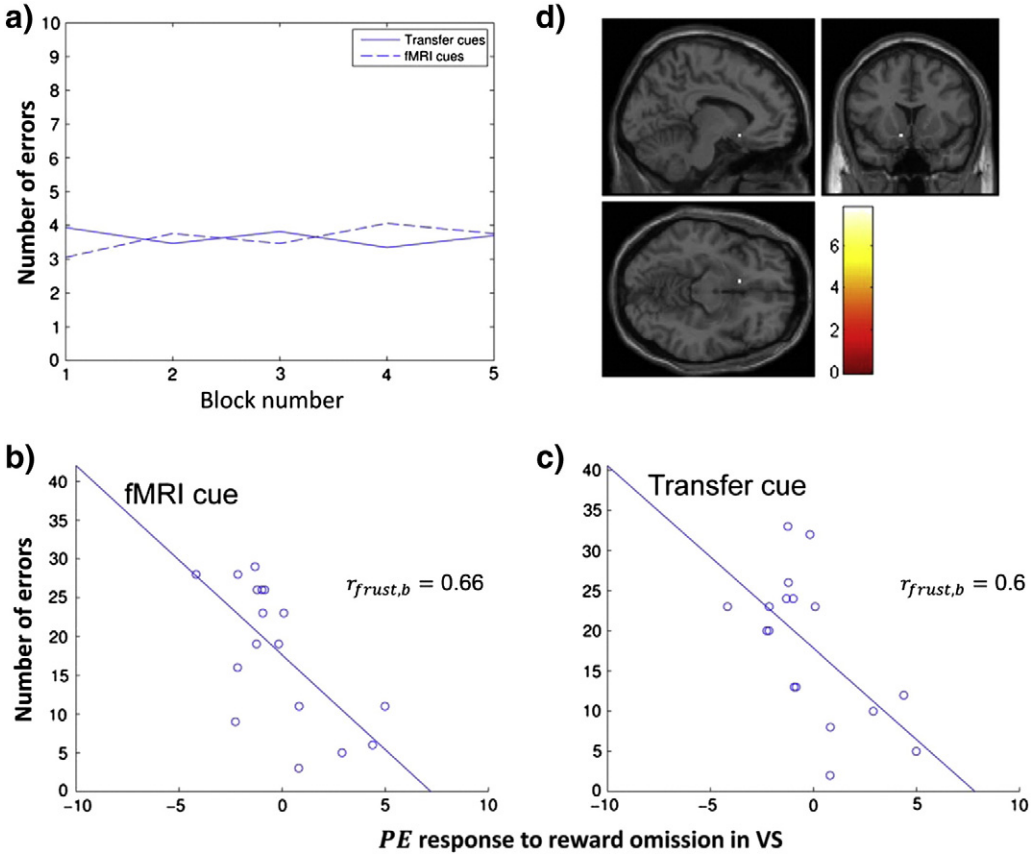

Fig. 3. Neuronal responses during associative learning predicted behavioral approach in follow-up session at a later date. a. Group average behavioral error rates are constant within and across the two phases. b. More negative ventral striatum responses to reward omission during passive associative learning predicted more costly behavioral choices at follow-up, i.e. more approach behavior towards cue A. b. This neuronal response also predicted more costly approach towards cue A′, encountered for the first time at follow-up. d. A whole brain analysis pointed to anatomical specificity: ventral striatum alone survived peak-level whole-brain correction.

cue, in our setting it can equivalently be understood as a model for learning the transition probabilities from cue to any outcome *state* (e.g. a fractal). Hence we can apply the same model to both the reward associations acquired by cue A and the non-reward associations of cue B. We distinguish these two learners by means of a superscript η_t^r for the erotic reward-predictive cue A, and the superscript η_t^s for the fractal-predictive cue B. The presence or absence of the outcome on any one trial was coded as $x = 1$ and $x = 0$, respectively. Rather than fitting the learning-rate parameter α – which required the assumption that there is only one learning rate in all brain regions – we performed 4 analyses, each assuming a different setting for α and covering the range of plausible settings of α i.e. 0.2, 0.4, 0.6, and 0.8.

Each analysis proceeded as follows. We trained η_t^r and η_t^s on the same trial-by-trial stimuli presented to each subject on the visual display as they lay in the MRI. We then introduced 2 columns in the each subject's first level fMRI design matrix as parametric modulators locked to the time of feedback. These predictors contained the positive/negative reward prediction errors which govern the dynamics of η_t^r, cue A's reward association. We analogously introduced two columns to encode positive/negative prediction errors underpinning the dynamics of η_t^s, cue B's fractal association. These four predictors may be written as δ^r_+,δ^r_-,δ^s_+,δ^s_- where, for example, the column δ^r_+ represents the value of the positive reward prediction error on any trial in which $\delta^r > 0$, and is 0 otherwise. Taken together, these four predictors specify the magnitude of a prediction error for every trial of the fMRI session. The design matrix also included predictors of no interest coding subjects' head movement. Estimating this model gave us the linear combination of δ^r_+,δ^r_-,δ^s_+,δ^s_- which best fits the fMRI data to each voxel × subject. We use the vector $\boldsymbol{\beta}$ to denote these best-fitting parameters.

We asked whether negative reward prediction errors δ^r_- versus negative state prediction errors δ^s_- successfully predicted subjects' error rate at behavioral follow-up. This quantity was estimated for each voxel × subject via the contrast $\boldsymbol{c^T\beta} = \beta_{\delta^r_-} - \beta_{\delta^s_-}$ (where subscripts on β indicate the column of the design matrix to which each parameter corresponds). As previously mentioned, we used simple linear regression in a second level model to explain between-subject variation in $\boldsymbol{c^T\beta}$ in terms of behavioral error at follow-up.

Experiment II

Participants

All 49 male subjects (19–28 year, median age 21) gave informed consent and were invited to read the task instructions. 24 were assigned to Treatment 1; 25 to Treatment 2.

Behavioral tasks

The purpose of this study was to provide a between subject confirmation that Pavlovian conditioning had on average impact on financially-costly approach behavior. Participants' earnings were calculated and paid out at the end of the experiment. Subjects learned on average 10.65 CHF, in addition to the show-up fee of 15 CHF. For both treatments, the initial passive conditioning session was performed on the same day in the behavioral lab and not under fMRI. Treatment 1 – the associative treatment – was otherwise identical to the experiment described in Figs. 1 and 2 while Treatment 2 – the non-associative treatment – did not involve associative learning during any passive trials in any session (only the erotic/fractal outcomes were presented). This ensured that non-specific sexual arousal was the same while the possibility of associative learning was eliminated.

Awareness measure

Subjects used the mouse to rate on a continuous scale, whether cue A or cue B was more associated with women in the passive 'rest' trials. Subjects also rated whether cue A or cue B was more associated with *points* in the active choice trials. These responses were transcribed onto the continuous[0,1] interval and coded such that a response of 0.5 indicated subjects had not detected any difference between the cues; responses closer to 1 indicated that subjects were correctly aware of the contingency; and responses closer to 0 indicated that subjects had a false belief about the contingency.

Results

Experiment I

How many choice errors did subjects commit in the active choice task? On average they chose the incorrect low-return cues A′ (in phase 1) and A (in phase 2) roughly 40% of the time (Fig. 3a), and the aggregate error rate does not differ between the original fMRI cues and the transfer cues (one sample *t*-test of the difference between the mean errors in fMRI and transfer cues, $p = 0.93$). Interestingly, the error rate also does not change much across the 5 blocks of each phase, indicating that subjects learned behavioral contingencies in the first block and their ability to make correct choices did not improve on average. For both phase 1 and phase 2 the difference between the first block and the fifth block was not significant (one sample *t*-test of the difference between the number of errors in the first versus the last block; phase 1: $p = 0.67$; phase 2: $p = 0.12$), indicating a stable aggregate error rate across blocks.

Model-independent analysis of prediction error

The aggregate error rate described above hides substantial individual differences in the proneness to make choice errors. We hypothesized that individuals with more potent associative learning under fMRI – indexed by greater RPE – would approach the low-return cues (A and A′) more often, sacrificing more money. To predict behavior from brain responses during passive associative learning, we correlated the number of times each subject chose cue A – their "behavioral error rate" in phase 2 – with their neuronal RPE, averaged within an anatomical mask of the ventral striatum. Fig. 3b shows that subjects with larger negative RPEs to (erotic) reward omission committed more behavioral errors ($p = 0.004$, $\rho = -0.66$, $n = 17$) during the five blocks of phase 2. Moreover, this correlation between the negative RPEs associated with erotic reward omission and choice errors prevails in all five blocks. In fact, the predictive power of RPEs even increased over the 5 blocks (the correlation in each block of phase 2 was: $\rho_1 = -0.43$, $\rho_2 = -0.58$, $\rho_3 = -0.62$, $\rho_4 = -0.63$, and $\rho_5 = -0.66$): it was separately significant at the $\alpha = 0.05$ level in all but the first block and was most significant in the final block ($p = 0.004$, $n = 17$; note that this p-value survives a conservative Bonferroni correction for 5 tests over the 5 separate blocks). This suggests that the initial learning signal during fMRI can predict subjects' asymptotic approach bias. Recall that cue A′ has formally identical contingencies to cue A, but is encountered for the first time at the beginning of the behavioral session. If a temporally stable *learning mechanism* is really at play, individuals who learnt potent Pavlovian associations to cue A during fMRI should more readily learn Pavlovian associations to the new cue A′, encountered for the first time at the behavioral session. By hypothesis, these subjects should therefore also come to approach A′ more often. We therefore asked if subjects' RPE during the associative learning task in the fMRI session with cue A could prospectively predict how readily a subject would acquire a financially-costly preference for cue A′ in the active choice task. To do this we correlated subjects' preference for the low-return gamble A′ with their RPE from the fMRI session and found a significant relationship, $p = 0.01$, $\rho = -0.6$, and $n = 17$ (Fig. 3c). The correlation in each block of phase 1 was: $\rho_1 = -0.35$, $\rho_2 = -0.57$, $\rho_3 = -0.62$, $\rho_4 = -0.61$, and $\rho_5 = -0.60$. This correlation was separately significant at the $\alpha = 0.05$ level in all but the first block. We therefore defined a single aggregate measure of each

subject's behavioral approach i.e. the number of A choices + the number of A′ choices.

We used simple linear regression to examine whether this statistical relation between aggregate behavioral measure and negative RPE existed in other brain regions. In a whole brain, between-subjects analysis we regressed RPE – the BOLD decrement following reward omission – on aggregate error. Using stringent peak-level FWE correction for multiple comparisons (Worsley et al., 1996) only one region, centered within the left ventral striatum, significantly correlated with this error at the 0.05 level: MNI coordinates (−9, 14, −11), see Fig. 3d.

Model-dependent analysis of prediction error

We observed a statistically significant relation between choice errors and negative prediction errors, as defined formally, at $p < 0.05$ set-level inference within a small volume mask of the ventral striatum, defined anatomically with a feature-inducing threshold of $p = 0.01$ uncorrected. This significant result held for learning rates α of 0.6 and 0.8, but not 0.2 and 0.4. This "set-level" effect carries the interpretation that there are more sub-regions within the ventral striatum displaying a relation between $\beta_{\delta^r_-} - \beta_{\delta^s_-}$ and choice error than would be expected by chance. We also examined whether the scale of positive reward prediction errors δ^r_+ versus positive state prediction errors δ^s_+ could successfully predict individual differences in behavioral choice error at follow-up. We found that positive reward prediction errors, relative to positive state prediction errors, i.e. $\beta_{\delta^r_+} - \beta_{\delta^s_+}$, significantly related to subjects' behavioral success at follow-up (again $p < 0.05$ set-level inference within a small volume mask of the ventral striatum, defined anatomically, as described above). This latter result held for a learning rate of 0.4, 0.6, and 0.8, but not 0.2.

It is noteworthy that these ventral striatum δ parameters have a weaker statistical relation to behavioral error than the more qualitative notion of negative prediction error used in the preceding analysis. That is, formally-defined δ parameters predict behavioral choice errors only at a more liberal statistical criterion: small-volume versus whole-brain corrected and at lower feature-inducing threshold corresponding to $p = 0.01$ uncorrected. In addition, we should strictly penalize this model-based analysis for the greater false-positive probability resulting from multiple comparisons over different settings of α. Conversely, as presented in the supplementary material, this model-based analysis revealed other regions outside of the ventral striatum whose δ activity correlated with behavioral choice error rate under stricter whole-brain correction. This contrasts with the preceding analysis, in which activation appeared more specific to the ventral striatum.

Cue-locked "reward anticipation" also predicts error

We also looked for regions in which BOLD responses to the presentation of the reward-predictive cue, rather than the outcome, can predict behavioral errors. Such responses can be interpreted as reward anticipatory responses that follow from prior learning. They are also predicted from models such as TD learning. In particular, this 'reward anticipatory' activation for each subject was defined as the difference in BOLD responses to the erotic-predictive CS versus the fractal-predictive CS. In a second-level analysis within the ventral striatum, defined anatomically, we observed three regions in which this anticipatory activation significantly predicted the number of behavioral errors (one in the left ventral striatum, two in the right ventral striatum). All exceeded both cluster and peak level significance at the level. The cluster inducing threshold required for the former inference was corresponded to $p = 0.001$ uncorrected. Increasing to a whole brain search, we found one highly significant, 304-voxel cluster in the dorsomedial prefrontal cortex centered at [3 35 49] in MNI coordinates, see supplementary Fig. 1.

Experiment II

We ran two behavioral control treatments to assess whether associative learning in particular, rather than some non-specific factor such as sexual arousal, was responsible for the behavioral approach bias observed at follow-up. These controls also examined whether subjects' awareness of the positive sexual associations influenced their bias. Control Treatment 1 ($n_1 = 24$) recapitulated the full fMRI/behavioral experiment above: prior passive training followed by intermingled passive and active blocks. Control Treatment 2 ($n_2 = 25$) was identical in every respect except that subjects never saw predictive cues A, A′, B, and B′ in any passive trial: They only saw erotic or fractals images in these trials (i.e. just the 'outcome'). Cues A, A′, B, and B′ still labeled the response options during active trials. Based on the fMRI study we had a plausible a priori hypothesis about the direction of the behavioral effect and therefore used one tailed tests for the behavioral control study. Without any positive associations for cue A in this control treatment, we expected less approach towards this cue during active choice trials. A two-sample one-sided *t*-test confirmed that on average subjects chose cue A less in the non-associative Treatment 2 than in the associative Treatment 1, $p = 0.022$ ($n_1 = 24, n_2 = 25$). Similarly, a two-sample one-sided *t*-test confirmed that subjects on average chose cue A′ less in the non-associative Treatment 2 than in the associative Treatment 1, $p = 0.043$ ($n_1 = 24, n_2 = 25$). The average error rate over blocks for cue A in the associative treatment was 42%, compared with 32% in the non-associative treatment. For cue A′ this was 41%, compared with 34%. Together with the analyses below, these results establish that, at the group level, associative learning influences approach behavior.

Time dynamics

We then asked whether the time dynamics of choice errors differed between the two treatment groups of the behavioral control study. To do this we used multilevel generalized linear regression to predict the number of errors in each 10-trial block in terms of the following predictors 1) Treatment condition — a binary indicator variable named '*associative*' which equal to 1 for the associative treatment and 0 for the non-associative treatment and 2) Time — a parametric variable named '*block*' ranging from 1,…,5 to indicate the progression of time over the 5 blocks. Critically we included the *associative* ∗ *block* interaction, which quantifies differences between the linear slopes in the associative versus non-associative treatment. Our statistical framework accommodates repeated-measures i.e. correlated choices within-subject (Gelman and Hill, 2007): it permits each subject their own intercept μ_i thereby modeling any unobserved subject-specific attribute which influences baseline error. Assuming these baseline error-rates μ_i are drawn from a Gaussian population distribution, this gives Eq. (1)

$$\begin{aligned} &P\left(Y_{ij} = y_{ij}\right) = Binomial\left(10, logit^{-1}\left(\eta_{ij}\right)\right) \\ &\eta_{ij} = \mu_i + \beta_1 block_{ij} + \beta_2 associative_{ij} + \beta_3 block_{ij} * associative_{ij} \\ &\mu_i \sim N\left(\mu, \sigma^2_\mu\right) \end{aligned} \quad (1)$$

where y_{ij} is the subject *i's* error rate in the *jth* block and ranges from 0, …,10, capital Y_{ij} is the corresponding random variable, $logit^{-1}$ is the inverse logistic function and $X \sim N(\mu,\sigma^2)$ denotes that X is distributed according to a Gaussian probability distribution with mean μ and variance σ^2.

In analogy to our preceding analyses, we estimated this model separately for choice errors made towards cue A and choice errors made towards "transfer cue" A′. There was no statistically significant difference between the linear time dynamics in the context of the "transfer cue" A′. In contrast, there was a statistically significant interaction in the context of cue A, indicating a more negative slope governing the dynamics of choice errors in the associative versus non-associative

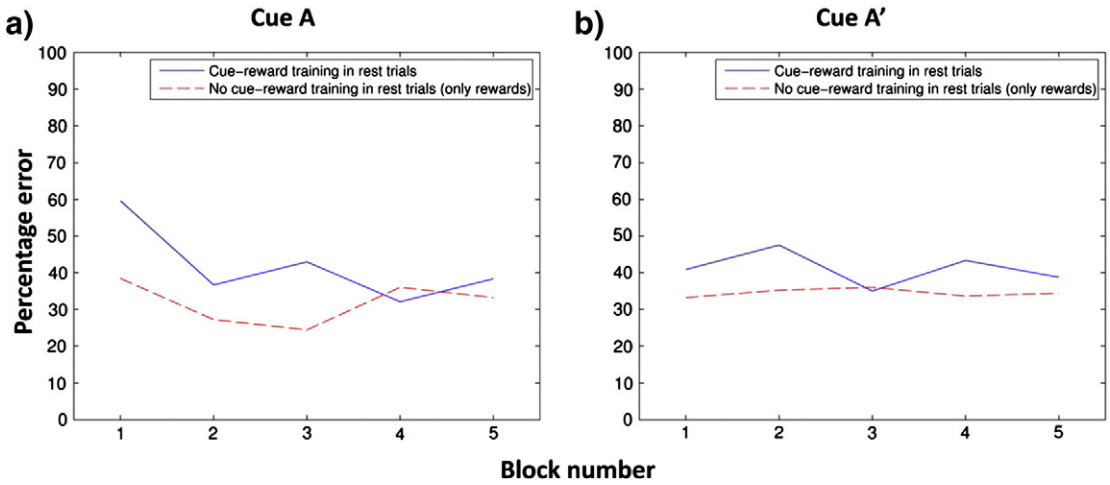

Fig. 4. Error rates in the behavioral control study. Time dynamics of error across the five 10-trial blocks for the non-associative treatment (dashed lines) — when there were rewards during rest trials but no predictive cues — and associative treatment (solid lines), when cues and rewards were paired during rest trials. a. Subjects chose cue A less on average in the non-associative treatment than in the associative treatment. In addition there was a significant difference in the slope of errors, with the associative treatment starting at a higher error rate and declining to a level comparable to the non-control group. See the main text for the corresponding statistical tests and interpretation. b. Subjects similarly chose cue A′ less on average in the non-associative treatment, but there was no significant time trend.

treatment, see also Fig. 4a ($p = 0.0011$, $n = n_1 + n_2 = 49$). The gradual extinction effect suggests that subjects were motivated to learn.

Awareness

After the experiment, we examined subjects' awareness of the passive and choice contingencies. A one-sided *t*-test confirmed that subjects were correctly aware of the choice contingency ($p = 0.0023$, $n_1 = 24$) and the cue-reward contingency ($p = 4 \times 10^{-8}$, $n_1 = 24$). Interestingly, unlike the RPE learning signal extracted from the fMRI, approach behavior could not be predicted from subjects' correct awareness that cue A had erotic associations (the correlation was insignificant at $p = 0.42$, $n_1 = 24$). This suggests that awareness of a cue's positive associations does not protect an individual from approaching it at cost.

Discussion

Simple Pavlovian learning does not provide a full account of even animal choice learning and is often viewed as just one factor influencing behavior (Seward, 1949; Thorpe, 1956; Dayan and Balleine, 2002; Dayan et al., 2006; Talmi et al., 2008; Guitart-Masip et al., 2011; Prévost et al., 2012). More sophisticated learning capabilities, each with their own quirks, underpin our habitual and goal-directed actions: these systems may not even require "prediction error" mechanisms. Given human cognitive sophistication, a key question is whether simple Pavlovian mechanisms influence human choice at all. We showed that individual differences are important. Specifically, functional variation between subjects in the ventral striatum as they learned to predict erotic events could explain a stable tendency to acquire and reacquire financially-costly behavior. Because this behavior persisted despite the consequences, it resembled the myopic, 'model-free' Pavlovian behavior observed in animal feeding (Domjan et al., 2000; Locurto, 1981; Rosenthal and Matthews, 1978; Sanabria et al., 2006; Williams and Williams, 1969), sexual and social behavior (Domjan et al., 2000). Our results therefore revive one classical perspective on the neurobiology of suboptimal decision-making: simple instinctive heuristics like 'approach things which predict reward', which were adaptive *on average* during natural selection, can easily leave us vulnerable to exploitation. Interestingly, subjects' awareness of a cue's misleading positive associations did not protect them from exploitation, echoing the claim that much of human choice behavior is under automatic stimulus control (Bargh and Chartrand, 1999).

Our task resembles both 'Pavlovian–instrumental-transfer' (PIT) paradigms and 'negative auto-maintenance' paradigms (Balleine, 2005; Guitart-Masip et al., 2010; Prévost et al., 2012; Sanabria et al., 2006; Talmi et al., 2008). The former paradigm entails separately training a passive Pavlovian cue–reward association and an instrumental action–reward association, then comparing responses in the presence versus absence of the 'irrelevant' Pavlovian cue in extinction. In the latter paradigm, Pavlovian and instrumental learning unfold concurrently in direct competition: a Pavlovian stimulus predicts reward only if the subject does not express a Pavlovian response. Our paradigm combines elements of both of these paradigms.

The main focus within our fMRI analysis is a quantity we refer to as the reward prediction error (RPE). More specifically our 'model-independent' analysis looked at a BOLD reduction in response to the omission of a likely reward. This is related to a RPE, but doesn't fully capture the RPE idea which varies over time (i.e. implements trial-by-trial learning) and also occurs on receipt of a reward that exceeds expectation, i.e. positive prediction errors. While positive prediction errors – e.g. as defined by Rescorla–×Wagner or TD learning – certainly occur in our task, they must be smaller in magnitude. This is a consequence of the contingencies in our design. Specifically, Rescorla–Wagner predicts that if a subject has learned that a reward will occur with 50% probability, then positive and negative prediction errors elicited on each trial should have the same magnitude, on average. Conversely if subjects expect reward with a probability higher than 50% – as in our case of 70% – then negative prediction errors will be larger than positive prediction errors. Thus our paradigm has more statistical sensitivity to identify effects in the domain of reward-omission, i.e. negative prediction errors. Additionally, there is not a satisfactory model-free way to quantify positive reward prediction errors in tasks such as ours. This is because positive reward prediction errors at the time of the outcome are typically confounded with the presence of sensory features of the outcome. For these reasons our model-independent analysis restricted attention to negative prediction errors, despite this not capturing all the qualities of the RPE. Interestingly, a model-dependent definition of prediction error yielded weaker correlations with choice error than this simple qualitative definition.

Future work might benefit from measuring subjects' sensitivity to the unconditioned erotic rewards (for example by measuring arousal/affective reactions). Individual differences in this sensitivity might ultimately explain variation in the strength of Pavlovian bias over subjects. For example an individual with greater reward sensitivity should, in theory, learn differently: In reward prediction error theories, reward sensitivity should directly control the asymptotic value assigned to

predictive cues, and therefore the magnitude of evoked prediction error as we measured it. The measurement of conditioned response amplitude (e.g. pupil reactions or another measure of conditioned arousal) would similarly augment the picture we present here. Larger Pavlovian bias should be expressed in higher amplitudes of these conditioned responses. Previous literature has repeatedly shown that the ventral striatum tracks Pavlovian learning effects (e.g. LaBar and Phelps, 2005; O'Doherty et al., 2004). We have therefore used responses in the ventral striatum as a proxy for Pavlovian learning. However, we should acknowledge that independent evidence of Pavlovian conditioning – e.g. changes in affective ratings for the cues, pupilometry or GSR – would bolster this interpretation. Relying on the ventral striatum activity as a means of inferring Pavlovian learning may be reasonably criticized if such activity reflects functions other than learning per se. The absence of independent-autonomic or behavioral-evidence for Pavlovian learning is therefore a noteworthy limitation in our experimental design. Further studies with appropriate behavioral measures will be needed to further support the claim that Pavlovian associations are underpinning the observed effects.

We have interpreted the relation between the ventral striatal BOLD responses and financially-costly choice in terms of a striatal learning mechanism which causes approach behavior i.e. model-free conditioned responses. It nonetheless remains possible that the ventral striatum predicts financially-costly choice in terms of a goal-directed or 'model-based', not a 'model-free', mechanism as follows. The ventral striatum responses reflect the acquired value of cue A, then subjects who value cue A more select it more at follow-up with the *nonpecuniary* objective or goal of magnifying it, because they like it. If this were true, then the relation between the ventral striatum BOLD responses and financially-costly choice would be particularly robust under the following conditions: 1) if subjects were unmotivated or slow to learn the *financially* optimal choice and/or 2) if they understood the financial trade-off but were willing to accept lower remuneration in order to see cue A magnified. Future work might assess whether subjects value non-financial aspects of the outcome. For example subjects may derive pleasure from viewing cue A at high magnification per se, because of its conditioned value. If they are willing to pay/forgo money in order to view cue A at high magnification, their behavior might be rationalized, i.e. might not be erroneous. The fact that subjects were explicitly aware of the instrumental contingencies in the post-hoc debriefing suggests that their choice was biased over and above instrumental learning of the financial costs.

Our control experiment demonstrates that in the absence of any "Pavlovian" biases, subjects choose the wrong cue 30% of the time, compared with 40% in the presence of Pavlovian biases. This implies that choice errors due to Pavlovian influences are, on average, around 10%. We will next briefly discuss potential explanations for high baseline errors, namely that some subjects did not understand the task or did not care about the rewards.

One possibility is that subjects erroneously chose the low value cue because they erroneously expected to gain information about its value. In fact, our task was designed to ensure that exploration was futile. Subjects were instructed "You will have two options. You can earn one point by selecting the correct option. Otherwise you will receive zero points", see supplementary Fig. 4. They therefore knew that there was exactly one correct response on each trial and could always infer the counterfactual outcome without resorting to exploration. This design feature was intended to exclude the alternative hypothesis that the ventral striatum in fact predicted individual differences in the preference for exploration versus exploitation (O'Doherty et al., 2004; Wittmann et al., 2007). It is possible that, despite this instruction, some subjects continued to explore the less valuable option under the false belief that they were gaining information about its value that could not be derived counterfactually. This misunderstanding would not be specific to any one of our experimental conditions, and could potentially explain high baseline errors throughout. Importantly however, we see no reason to suspect that individual differences in the ventral striatal activation predicted individual differences in the incomprehension of this one feature of the task instructions. Thus, it does not provide a compelling alternative explanation for our main fMRI results.

Another possibility is that high baseline errors are an expression of the matching law (Killeen, 1972; Poling et al., 2011), which states that the choice rate (for cue A versus cue B) equals the relative reinforcement rate for these two cues. However, as reported in the supplementary material, we did not find evidence that subjects' behavior respected the matching law. Finally, some subjects may simply have cared less about financial outcomes, which would provide an alternative explanation for errors. This opens the possibility that those subjects showing larger fMRI omission responses during Pavlovian conditioning also care less about the financial outcomes and were less motivated to learn the correct choice. If this were true, then the increased error rate in these individuals reflects greater randomness in their behavior rather than a specific targeted approach of cue A. In general it is difficult in our paradigm to dissociate subtle changes in approach behavior from subtle changes in the randomness of subject's choice behavior. Specifically, any change which pushes a subject away from deterministic, money-maximizing behavior (choosing cue B deterministically) but not as far as deterministic approach behavior towards cue A may be difficult to dissociate from simple increases in randomness. This difficulty mirrors the difficulty encountered when trying to distinguish increases in exploration from increases in randomness due to other factors.

Widespread practical consequences would follow from the conclusion that human choice behavior is under automatic stimulus control. For example, recent behavioral work appears to implicate unconditioned Pavlovian responses in consumer behavior (Bushong et al., 2010). From our perspective, it is relevant that advertisements often pair some rewarding (e.g. erotic) stimulus (Gresham and Shimp, 1985) with a product whose consumption contradicts our interests and goals. Our results also generate new hypotheses about defective decision-making in the clinic. For example, 'behavioral' addictions or 'compulsions' (Parashar and Varma, 2007; Potenza, 2006; Stein et al., 2010) often involve a self-punitive appetite for sex, pornography, gambling/gaming or food which persist despite excessive physical, mental, social and/or financial consequences. While ventral striatal DA has been implicated in virtually all stages of drug addiction, including associatively 'cued' drug consumption (Parashar and Varma, 2007), we are largely ignorant of the underlying pathogenesis in non-substance-related addictions. Our results suggest that a ventral striatal DA *learning mechanism* may also help explain these conditions, suggesting why vulnerable individuals tend to (re)acquire addictive dispositions to multiple behaviors and substances (Cunningham-Williams et al., 1998; Petry et al., 2005).

Acknowledgments

We would like to thank Peter Dayan and two anonymous reviewers for helpful comments on this manuscript. JRC was funded by SystemsX, Switzerland.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.11.039.

References

Balleine, B., Killcross, S., 1994. Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. Behav. Brain Res. 65 (2), 181–193.
Balleine, B.W., 2005. Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol. Behav. 86 (5), 717–730.
Balleine, B.W., O'Doherty, J.P., 2009. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35 (1), 48–69.

Bargh, J.A., Chartrand, T.L., 1999. The unbearable automaticity of being. Am. Psychol. 54 (7), 462.
Berridge, K.C., 2007. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology 191 (3), 391–431.
Berridge, K.C., Robinson, T.E., 1998. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res. Rev. 28 (3), 309–369.
Boureau, Y.L., Dayan, P., 2010. Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36 (1), 74–97.
Bushong, B., King, L.M., Camerer, C.F., Rangel, A., 2010. Pavlovian processes in consumer choice: the physical presence of a good increases willingness-to-pay. Am. Econ. Rev. 100 (4), 1556–1571.
Corbit, L.H., Balleine, B.W., 2003. The role of prelimbic cortex in instrumental conditioning. Behav. Brain Res. 146 (1–2), 145–157.
Corbit, L.H., Muir, J.L., Balleine, B.W., 2001. The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. J. Neurosci. 21 (9), 3251–3260.
Cunningham-Williams, R.M., Cottler, L.B., Compton, W.M., Spitznagel, E.L., 1998. Taking chances: problem gamblers and mental health disorders–results from the St. Louis Epidemiologic Catchment Area Study. Am. J. Public Health 88 (7), 1093–1096.
Dayan, P., Balleine, B.W., 2002. Reward, motivation, and reinforcement learning. Neuron 36 (2), 285–298.
Dayan, P., Niv, Y., Seymour, B., Daw, N.D., 2006. The misbehavior of value and the discipline of the will. Neural Netw. 19 (8), 1153–1160.
Domjan, M., Cusato, B., Villarreal, R., 2000. Pavlovian feed-forward mechanisms in the control of social behavior. Behav. Brain Sci. 23 (02), 235–249.
Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. Q. J. Econ. 75 (4), 643–669.
Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
Gresham, L.G., Shimp, T.A., 1985. Attitude toward the advertisement and brand attitudes: a classical conditioning perspective. J. Advert. 14 (1), 10–17.
Guitart-Masip, M., Fuentemilla, L., Bach, D.R., Huys, Q.J., Dayan, P., Dolan, R.J., Duzel, E., 2011. Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J. Neurosci. 31 (21), 7867–7875.
Guitart-Masip, M., Talmi, D., Dolan, R., 2010. Conditioned associations and economic decision biases. NeuroImage 53 (1), 206–214.
Hall, J., Parkinson, J.A., Connor, T.M., Dickinson, A., Everitt, B.J., 2001. Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. Eur. J. Neurosci. 13 (10), 1984–1992.
Ikemoto, S., Panksepp, J., 1999. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res. Rev. 31 (1), 6–41.
Killcross, S., Robbins, T.W., Everitt, B.J., 1997. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature 388 (6640), 377–380.
Killeen, P., 1972. The matching law. J. Exp. Anal. Behav. 17 (3), 489–495.
LaBar, K.S., Phelps, E.A., 2005. Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. Behav. Neurosci. 119 (3), 677–686.
Locurto, C., 1981. Contributions of autoshaping to the partitioning of conditioned behavior. Autoshaping Cond. theory 101–135.
Mogenson, G.J., Jones, D.L., Yim, C.Y., 1980. From motivation to action: functional interface between the limbic system and the motor system. Prog. Neurobiol. 14 (2–3), 69–97.
O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J., 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304 (5669), 452–454.
O'Doherty, J.P., Buchanan, T.W., Seymour, B., Dolan, R.J., 2006. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. Neuron 49 (1), 157.
Panksepp, J., 2004. Affective Neuroscience: The foundations of Human and Animal Emotions. Oxford University Press, USA.
Parashar, A., Varma, A., 2007. Behavior and substance addictions: is the world ready for a new category in the DSM-V? CNS Spectr. 12 (4), 257 (author reply 258).
Parkinson, J.A., Willoughby, P.J., Robbins, T.W., Everitt, B.J., 2000. Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: further evidence for limbic cortical–ventral striatopallidal systems. Behav. Neurosci. 114 (1), 42.
Petry, N.M., Stinson, F.S., Grant, B.F., 2005. Comorbidity of DSM-IV pathological gambling and other psychiatric disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J. Clin. Psychiatry 66 (5), 564–574.
Poling, A., Edwards, T.L., Weeden, M.A., Foster, T.M., 2011. The Matching Law.
Potenza, M.N., 2006. Should addictive disorders include non-substance-related conditions? Addiction 101 (s1), 142–151.
Prévost, C., Liljeholm, M., Tyszka, J.M., O'Doherty, J.P., 2012. Neural correlates of specific and general Pavlovian-to-instrumental transfer within human amygdalar subregions: a high-resolution fMRI study. J. Neurosci. 32 (24), 8383–8390.
Reading, P.J., Dunnett, S.B., Robbins, T.W., 1991. Dissociable roles of the ventral, medial and lateral striatum on the acquisition and performance of a complex visual stimulus-response habit. Behav. Brain Res. 45 (2), 147–161.
Rescorla, R., Wagner, A., 1972. Variations in the Effectiveness of Reinforcement and Nonreinforcement. Classical Conditioning II: Current Research and Theory, Appleton-Century-Crofts, New York.
Reynolds, S.M., Berridge, K.C., 2001. Fear and feeding in the nucleus accumbens shell: rostrocaudal segregation of GABA-elicited defensive behavior versus eating behavior. J. Neurosci. 21 (9), 3261.
Robbins, T., Giardini, V., Jones, G., Reading, P., Sahakian, B., 1990. Effects of dopamine depletion from the caudate–putamen and nucleus accumbens septi on the acquisition and performance of a conditional discrimination task. Behav. Brain Res. 38 (3), 243–261.
Rosenthal, R.L., Matthews, T.J., 1978. The effects of prefeeding in autoshaping and omission training. Bull. Psychon. Soc. 11 (3), 153–156.
Salamone, J.D., Correa, M., 2002. Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. Behav. Brain Res. 137 (1–2), 3–25.
Sanabria, F., Sitomer, M.T., Killeen, P.R., 2006. Negative automaintenance omission training is effective. J. Exp. Anal. Behav. 86 (1), 1.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599.
Sesack, S.R., Grace, A.A., 2009. Cortico-basal ganglia reward network: microcircuitry. Neuropsychopharmacology 35 (1), 27–47.
Seward, J.P., 1949. An experimental analysis of latent learning. J. Exp. Psychol. 39 (2), 177.
Stein, D.J., Hollander, E., Rothbaum, B.O., 2010. Textbook of Anxiety Disorders. Amer Psychiatric Pub Incorporated.
Talmi, D., Seymour, B., Dayan, P., Dolan, R.J., 2008. Human Pavlovian–instrumental transfer. J. Neurosci. 28 (2), 360.
Thorpe, W.H., 1956. Learning and Instinct in Animals.
Tobler, P.N., Fletcher, P.C., Bullmore, E.T., Schultz, W., 2007. Learning-related human brain activations reflecting individual finances. Neuron 54 (1), 167–175.
Williams, D.R., Williams, H., 1969. Auto-maintenance in the pigeon: sustained pecking despite contingent non-reinforcement. J. Exp. Anal. Behav. 12 (4), 511.
Wittmann, B.C., Bunzeck, N., Dolan, R.J., Duzel, E., 2007. Anticipation of novelty recruits reward system and hippocampus while promoting recollection. Neuroimage 38 (1), 194–202.
Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4 (1), 58–73.