Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding

Quentin J. M. Huys[1,2,3]*, Roshan Cools[4], Martin Gölzer[5], Eva Friedel[5], Andreas Heinz[5], Raymond J. Dolan[1], Peter Dayan[2]

[1] Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom, [2] Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom, [3] Medical School, University College London, London, United Kingdom, [4] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Nijmegen, Netherlands, [5] Charité Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany

Abstract

Hard-wired, Pavlovian, responses elicited by predictions of rewards and punishments exert significant benevolent and malevolent influences over instrumentally-appropriate actions. These influences come in two main groups, defined along anatomical, pharmacological, behavioural and functional lines. Investigations of the influences have so far concentrated on the groups as a whole; here we take the critical step of looking inside each group, using a detailed reinforcement learning model to distinguish effects to do with value, specific actions, and general activation or inhibition. We show a high degree of sophistication in Pavlovian influences, with appetitive Pavlovian stimuli specifically promoting approach and inhibiting withdrawal, and aversive Pavlovian stimuli promoting withdrawal and inhibiting approach. These influences account for differences in the instrumental performance of approach and withdrawal behaviours. Finally, although losses are as informative as gains, we find that subjects neglect losses in their instrumental learning. Our findings argue for a view of the Pavlovian system as a constraint or prior, facilitating learning by alleviating computational costs that come with increased flexibility.

Citation: Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, et al. (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. PLoS Comput Biol 7(4): e1002028. doi:10.1371/journal.pcbi.1002028

Editor: Antonio Rangel, California Insitute of Technology, United States of America

Received November 12, 2010; **Accepted** February 22, 2011; **Published** April 21, 2011

Funding: This work was supported by a Wellcome Trust Grant and a Max Planck Award to RJD. PD was supported by the Gatsby Charitable Foundation. Publication costs were shared between the Max Planck Award and the Gatsby Charitable Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: qhuys@cantab.net

Introduction

The functional architecture of responding involves two fundamental components that are behaviourally [1] and computationally [2] separable: Pavlovian and instrumental. The instrumental component respects the stimulus-dependent contingency between responses and their outcomes (stimulus-response and action-outcome learning) [3]. By contrast, preparatory Pavlovian responses, chiefly involving approach and withdrawal, are elicited by the appetitive or aversive valence associated with predictive stimuli in a manner that is not dependent on the consequences of those responses [3–5].

The interactions between the two systems are most evident when automatically-elicited Pavlovian responses interfere with contingent instrumental responding [1,6–9]. For instance, pigeons will strikingly continue to peck at a light predictive of food (a preparatory approach elicited by the appetitive prediction), even if the food is withheld every time they peck the light (the instrumental contingency) [10,11]. Pavlovian interference likely contributes to many quirks of behaviour such as impulsivity [12], framing and [13], endowment effects [14] and many other "anomalies" [15], including neurological [16–19] and psychiatric diseases [20–26]. Further, puzzling facets of seemingly purely instrumental behaviour such as the difficulties in learning 'go' responses to avoid punishments; or 'nogo' to obtain rewards (unpublished data) and even the restrictions in associations evident in 'evolutionarily preparedness' [27,28] might be traced to Pavlovian principles.

However, instrumental and Pavlovian systems share overlapping neural hardware. Their bidirectional interaction is characterised by two key triads: rewards are tied to approach and vigour; and punishments to withdrawal and behavioural inhibition. The neuromodulator dopamine (DA) responds predominantly to rewards [22,29–31], induces behavioural activation and enhances approach [32–35]. Each aspect of this triad confounds the role of the phasic DA bursts in the flexible acquisition of instrumental values [36–42]. Serotonin appears to lie at the heart of the aversive triad, having been linked to punishments [43–45], behavioural inhibition and withdrawal [25,32,46–52], although dopamine acting via D2 receptors likely also plays a role in linking absence of rewards to nogo [17,53,54]. Signatures of both triads are also evident in neural circuits involved in response and choice. In the dorsal striatum, there are interdigitated pathways for 'go' and 'nogo', with the go pathways again linked positively to rewards via dopamine [16,18,55,56]. The ventral striatum is primarily organized along an appetitive/aversive axis with direct links to approach and withdrawal behaviours [57,58]. The aversive triad is also tightly linked to the dorsal raphé and the periaquaeductal gray [59,60].

The main routes to the scientific investigation of these interactions consists of tasks in which Pavlovian stimuli are presented during ongoing instrumental tasks. However, these have

Author Summary

Beautiful background music in a shop may well tempt us to buy something we neither need nor want. Valenced stimuli have broad and profound influences on ongoing choice behaviour. After replicating known findings whereby approach is enhanced by appetitive Pavlovian stimuli and inhibited by aversive ones, we extend this to withdrawal behaviours, but critically controlling for the valence of the withdrawal behaviours themselves. We find that even when withdrawal is appetitively motivated, it is still inhibited by appetitive Pavlovian stimuli and enhanced by aversive ones. This shows, for the first time, that the effect of background Pavlovian stimuli depends critically on the intrinsic valence of behaviours, and differs between approach and withdrawal.

as yet not explored the full set of interactions characterising the overlap between the two systems. Two critical confounds remain: The first confound concerns the precise nature of the effect of Pavlovian stimuli on instrumental behaviours. The instrumental behaviours studied have largely been appetitively motivated approach behaviours (in Pavlovian-Instrumental Transfer (PIT) and conditioned suppression tasks, [1,6–8,61–63]), and one instance of aversively motivated withdrawal behaviour [64]. The relative role of the appetitive-aversive motivation axis versus that of the approach-withdrawal axis is unknown. This in turn obscures the nature of the interaction: whether Pavlovian stimuli interact with the value of the instrumental behaviour, or by promoting specific responses [1], or even simply by modulating behavioural activation [5]. Second, the extent to which the separation of reward and punishment processing into opponent motivational structures applies to instrumental as well as Pavlovian learning is incompletely explored [1,27,28,65].

All these issues can simultaneously be addressed in a combined PIT and conditioned suppression task with both approach and withdrawal actions in which the overall motivational component of approach and withdrawal are matched (Figure 1 and Table 1). The task separates the contributions of approach and withdrawal by using two counterbalanced blocks, one involving approach go versus nogo, and the other withdrawal go versus nogo. The comparison between go and nogo controls for effects of behavioural activation or inhibition. In each block, subjects first underwent brief instrumental training (Figure 1A), learning from positive *and* negative feedback (monetary gains and losses of €0.20) whether to produce a go or a nogo response associated with sorting mushrooms. In the approach block (Figure 1A, top, all 46 subjects), go responses involved moving the cursor onto a mushroom (to collect it), while nogo involved doing nothing, thus not collecting the mushroom. To test for the effect of low-level motor variables, subjects performed one of two types of withdrawal actions. In "throwaway" (24 subjects, Figure 1A, middle), go involved moving the cursor physically away from the mushroom and clicking into an empty blue box; nogo involved doing nothing, and thus keeping the mushroom. Importantly, both approach to and withdrawal from the instrumental stimulus were orthogonal to any approach and withdrawal that might be directed at the Pavlovian background stimulus. In "release" (22 subjects,

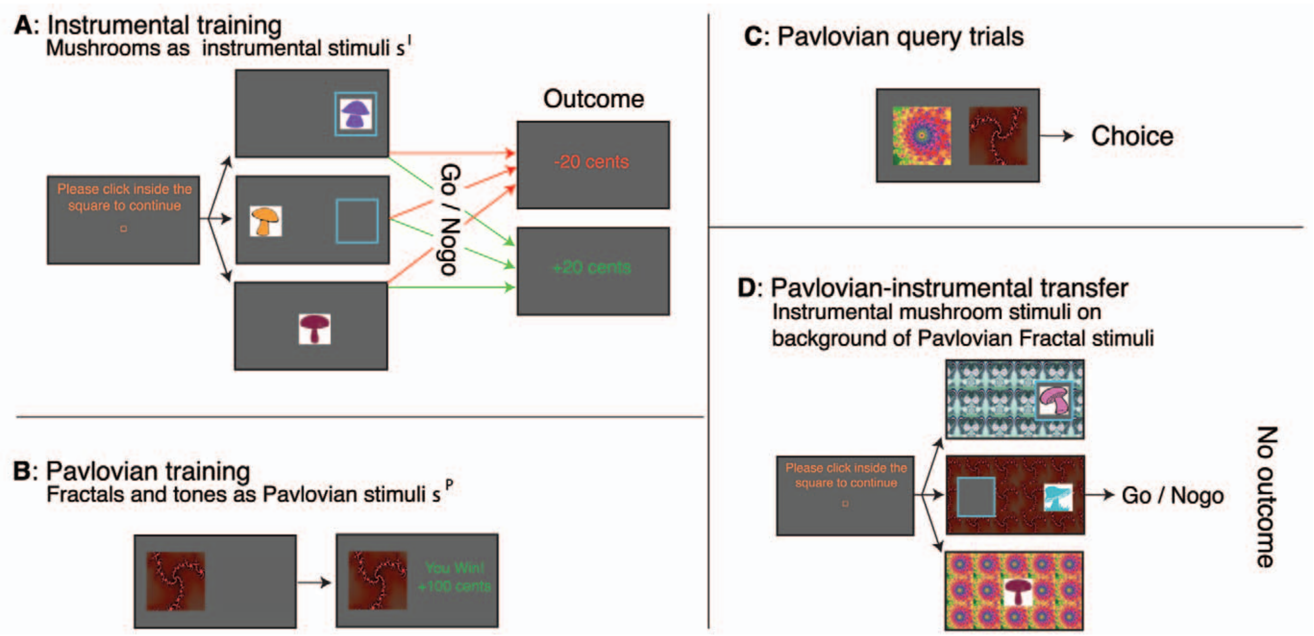
A: Instrumental training
Mushrooms as instrumental stimuli s^I
Please click inside the square to continue
Outcome
-20 cents
+20 cents
Go / Nogo
B: Pavlovian training
Fractals and tones as Pavlovian stimuli s^P
You Win!
+100 cents
C: Pavlovian query trials
Choice
D: Pavlovian-instrumental transfer
Instrumental mushroom stimuli on background of Pavlovian Fractal stimuli
Please click inside the square to continue
Go / Nogo
No outcome

Figure 1. Task description. A: Instrumental training. To centre the cursor, subjects clicked in a central square. In approach trials (top), subjects chose whether to move the cursor towards the mushroom and click inside the blue frame onto the mushroom (go), or not do anything (nogo). In throwaway withdrawal trials (middle), they instead moved the cursor away from the mushroom and clicked in the empty blue frame (go) or did nothing (nogo). In release withdrawal trials (bottom), subjects were instructed to keep the button pressed after the initial click in the central square. The mushroom was then presented centrally, under the cursor. To throw away the mushroom, subjects released the button. Outcomes were presented immediately after go actions, or after 1.5 seconds. **B**: Pavlovian training. Subjects passively viewed stimuli and heard auditory tones, followed by wins and losses. **C**: On Pavlovian query trials, subjects chose between two Pavlovian stimuli. No outcomes were presented, but they were counted and added to the total presented at the end of the experiment. **D**: Pavlovian-instrumental transfer. Subjects responded to instrumental stimuli with Pavlovian stimuli tiling the background. No outcomes were presented, but subjects were instructed that their choices counted towards the final total. No explicit instructions about the contribution of Pavlovian stimuli towards the final total were given.
doi:10.1371/journal.pcbi.1002028.g001

Table 1. Experimental layout.

Approach Block			
A1	Instrumental training (60 trials)		Probabilistic reinforcements[1]: ± 0.20 €
		$s^{\mathcal{I}}_{1,2,3} \rightarrow$approach	$p(\mathrm{rew}\vert\mathrm{go},s^{\mathcal{I}}_{1,2,3})=0.7$, $p(\mathrm{pun}\vert\mathrm{go},s^{\mathcal{I}}_{1,2,3})=0.3$[1]
		$s^{\mathcal{I}}_{4,5,6} \rightarrow$nogo	$p(\mathrm{rew}\vert\mathrm{nogo},s^{\mathcal{I}}_{4,5,6})=0.7$, $p(\mathrm{pun}\vert\mathrm{nogo},s^{\mathcal{I}}_{4,5,6})=0.3$[1]
A2	Pavlovian training (60 trials)		Deterministic reinforcements
		$s^{\mathcal{P}}_{++} \rightarrow$reward	1 €
		$s^{\mathcal{P}}_{+} \rightarrow$reward	0.10 €
		$s^{\mathcal{P}}_{0} \rightarrow$	0
		$s^{\mathcal{P}}_{-} \rightarrow$punishment	−0.10 €
		$s^{\mathcal{P}}_{--} \rightarrow$punishment	−1 €
A3	PIT (100 trials)		No Reinforcements
		$s^{\mathcal{P}} \times s^{\mathcal{I}}_{1-6} \rightarrow$?	
Withdrawal Block			
W1	Instrumental training (60 trials)		Probabilistic reinforcements[1]: ± 0.20 €
		$s^{\mathcal{I}}_{7,8,9} \rightarrow$withdraw	$p(\mathrm{rew}\vert\mathrm{go},s^{\mathcal{I}}_{7,8,9})=0.7$, $p(\mathrm{pun}\vert\mathrm{go},s^{\mathcal{I}}_{7,8,9})=0.3$[1]
		$s^{\mathcal{I}}_{10,11,12} \rightarrow$nogo	$p(\mathrm{rew}\vert\mathrm{nogo},s^{\mathcal{I}}_{10,11,12})=0.7$, $p(\mathrm{pun}\vert\mathrm{nogo},s^{\mathcal{I}}_{10,11,12})=0.3$[1]
W2	Pavlovian training (60 trials)		Deterministic reinforcements
		$s^{\mathcal{P}}_{++} \rightarrow$reward	1 €
		$s^{\mathcal{P}}_{+} \rightarrow$reward	0.10 €
		$s^{\mathcal{P}}_{0} \rightarrow$	0
		$s^{\mathcal{P}}_{-} \rightarrow$punishment	−0.10 €
		$s^{\mathcal{P}}_{--} \rightarrow$punishment	−1 €
W3	PIT (100 trials)		No Reinforcements
		$s^{\mathcal{P}} \times s^{\mathcal{I}}_{7-12} \rightarrow$?	

Note the numerical subscripts on the instrumental stimuli $s^{\mathcal{I}}$ here refer to their identities, not to the time of presentation.

[1] For subject with deterministic instrumental reinforcements, the outcome probabilities were 1 and 0 instead of 0.7 and 0.3, respectively.

doi:10.1371/journal.pcbi.1002028.t001

Figure 1A, bottom), the subjects had to start by pressing the mouse button. Go involved releasing the button to avoid collecting the mushroom; nogo involved continuing to press the button and thereby receiving the mushroom.

In order to orthogonalise the approach-withdrawal and appetitive-aversive axes, the *learned* instrumental values in approach and withdrawal blocks needed to be matched. To achieve this, both go and nogo responses were, if correct, rewarded. Additionally, to avoid the confound of activation, in each block (i.e. in both approach and withdrawal blocks) the go action was designated as the correct response to half the instrumental stimuli, and the nogo action to the other half (see Table 1). Incorrect responses had opposite outcome contingencies to correct responses, yielding more punishments than rewards. This ensured that go, nogo, approach and withdrawal overall had the same learned association with rewards and punishments. We tested both deterministic and probabilistic outcomes but found no differences.

In the second part of each block, subjects passively viewed unrelated, fractal, stimuli paired with separate auditory tones (Figure 1B). Each compound Pavlovian stimulus $s^{\mathcal{P}}$ was deterministically associated with a monetary gain or loss, i.e. its Pavlovian value $\mathcal{V}(s^{\mathcal{P}})$ was equal to that monetary outcome. Every fifth trial in the Pavlovian block was a query trial (Figure 1C), in which subjects chose the better of two fractal visual stimuli without being informed about the outcome. Finally, in the PIT stage, the instrumental stimuli were presented on a background of fractal Pavlovian stimuli together with the auditory tones, and again without outcome information.

Our task addressed the key confounds described above. With respect to the triads, we found that the Pavlovian influence is action specific: appetitive Pavlovian cues boosted go approach responses and suppressed withdrawal go responses; aversive Pavlovian cues did the opposite. Additionally, subjects were substantially biased against withdrawal, but we found no evidence that the instrumental learning component itself differed between the approach and withdrawal condition.

Results

The key results in this paper concern the interaction of valued Pavlovian stimuli on instrumental choices. We first present a direct analysis of the choice data and reaction times. We then provide a detailed modelling analysis of the data, employing a stringent form of group-level model selection that assesses each model's parsimony by weighing its ability to fit the data against its complexity. The models quantify Pavlovian values $\mathcal{V}(s^{\mathcal{P}})$, which are the expectations of a gain or loss given Pavlovian stimulus $s^{\mathcal{P}}$, and instrumental choice values $\mathcal{Q}_t(a,s^{\mathcal{I}})$, which are the time-varying expectations of a reward given a response a to an instrumental stimulus $s^{\mathcal{I}}$. The structure of the most parsimonious model implies the influences and interactions that were significant (for instance ruling in a bias against active withdrawal, but ruling out any difference between the instrumental learning rates associated with approach and withdrawal); the values of the parameters in this model indicate the nature of those influences and interactions.

Model-free analyses

There was no difference between the results for probabilistic and deterministic feedback, and we therefore present the combined data. Analysis of the components of the experiment indicate robust, yet moderate, instrumental conditioning that was stable during the PIT period, combined with highly robust Pavlovian conditioning. Figure 2A shows the instrumental probability of choosing the more rewarded ("correct") stimulus over time. Subjects rapidly came to prefer the more rewarded action. Preference was weaker for go withdrawal, against which there was a consistent bias. We intended the instrumental preference to be weak to avoid ceiling effects when assessing PIT.

Subjects also exhibited predictable variability on a shorter time-scale: Figure 2B shows the immediate consequences of rewards and punishments on subsequent behaviour. It is notable that punishments did not reduce the repeat probability below chance level (mean $p(\mathrm{switch}_t|\mathrm{pun}_{t-1})$ is not <0.5, one-tailed t-test $p>.2$). The same was found when analysing go and nogo choices separately: in both cases, $p(\mathrm{switch}_t|\mathrm{pun}_{t-1})$ was not significantly different from 0.5 (both $p>.3$, two-tailed t-test), and was significantly smaller than $p(\mathrm{stay}_t|\mathrm{rew}_{t-1})$ (both $p<4\times10^{-6}$, paired t-test). Whether this really does represent an insensitivity to punishments depends, however, on the average stay probability, and on how this average stay probability is related to past reinforcements. Subjects were instructed that the outcomes of responses in the PIT block would be counted as in the instrumental block. Figure 2C shows that this led to stable maintenance of the instrumental response tendencies throughout the PIT block. Figure 2D shows that all but one (excluded) subject showed extremely good performance on the Pavlovian query trials interleaved with the Pavlovian training (mean correct $>95\%$).

Given the success of instrumental and Pavlovian training, we next analysed the raw effect of Pavlovian stimuli on approach and withdrawal choices. Figure 2E shows a highly significant interaction between block and Pavlovian stimulus valence. Relative to neutral stimuli, positive Pavlovian stimuli enhanced approach and inhibited withdrawal go over nogo. Conversely, negative Pavlovian stimuli enhanced withdrawal and inhibited approach go over nogo. A similar analysis looking at the probability of responding incorrectly (outside the blue box) showed no effect of the Pavlovian stimuli in either approach or withdrawal condition and no interaction ($p=0.26,0.22,0.88$ respectively, ANOVA), suggesting that these results were not due to response competition. Note that the withdrawal go probabilities were lower than the approach ones, again reflecting the overall bias against go withdrawal.

Average reaction times for go approach and go withdrawal actions did not differ ($p=0.097$, 2-tailed t-test). Against our expectations, Pavlovian stimuli of both positive and negative valence shortened reaction times in a parametric manner relative to neutral Pavlovian stimuli (Figure 2F, $p=0.0310$, ANOVA), although this effect was not present in either block separately ($p=0.5502$ and $p=0.0781$ respectively, ANOVA).

Model-based analyses

The size of the PIT effect may have been affected by the extent of instrumental learning (and thus the actual learned action values), by response biases, and by generalization from the instrumental to the PIT stage. In addition, there may have been differences in the instrumental learning of approach and withdrawal actions (Figure 2A). We decomposed and analysed all such factors using a detailed reinforcement learning model. This contained explicit parameters capturing all the instrumental and Pavlovian effects in

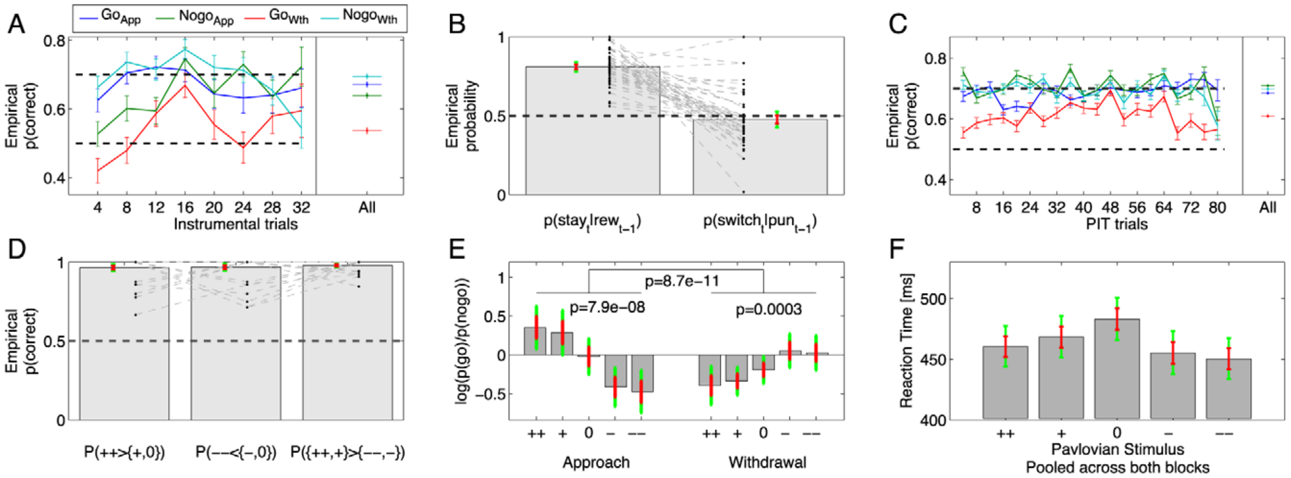

Figure 2. Raw choice probabilities. A&C: Average probability (±1 standard error) of choosing the more rewarded ("correct") action in the instrumental (A) and PIT (C) parts. Average performance was above chance in all cases, but worse when withdrawal go was the more rewarded action (red). There was no extinction during the PIT block. Each point is the average across subjects and across four trials. **B**: The bars show mean overall probability of repeating an action in the instrumental part given that it was last rewarded in the presence of the current stimulus, or the probability of switching given a previous punishment. Punishments do not lead to reliable switching. **D**: Choice probabilities in the Pavlovian forced choice query trials. Most subjects were close to perfect. The grey bars show the probabilities of *left*: choosing a very good stimulus (++) over a good (+) or neutral (0) stimulus; *middle*: choosing a bad (−) or neutral (0) stimulus over a very bad (--) stimulus; *right*: choosing a positive (++ or +) stimulus over a negative one (-- or -). Subjects that performed submaximally in the appetitive Pavlovian domain did not necessarily have lower reward sensitivities in the instrumental task, and vice versa for aversive Pavlovian stimuli and punishment sensitivity. **E**: PIT effects. The left part shows the approach PIT block, the right part the withdrawal PIT block. Each bar shows the log ratio of the choice probability (go/nogo) in the presence of one of the five Pavlovian stimuli. There was a significant effect of Pavlovian stimulus valence in each block. In addition, there was a significant block × Pavlovian stimulus valence interaction. Grey bars are means ±1 standard error (red) and $\pm95\%$ confidence intervals (green). **F**: Reaction times, pooled data for both PIT blocks. The bigger the absolute valence of the Pavlovian stimulus, the shorter the reaction time.
doi:10.1371/journal.pcbi.1002028.g002

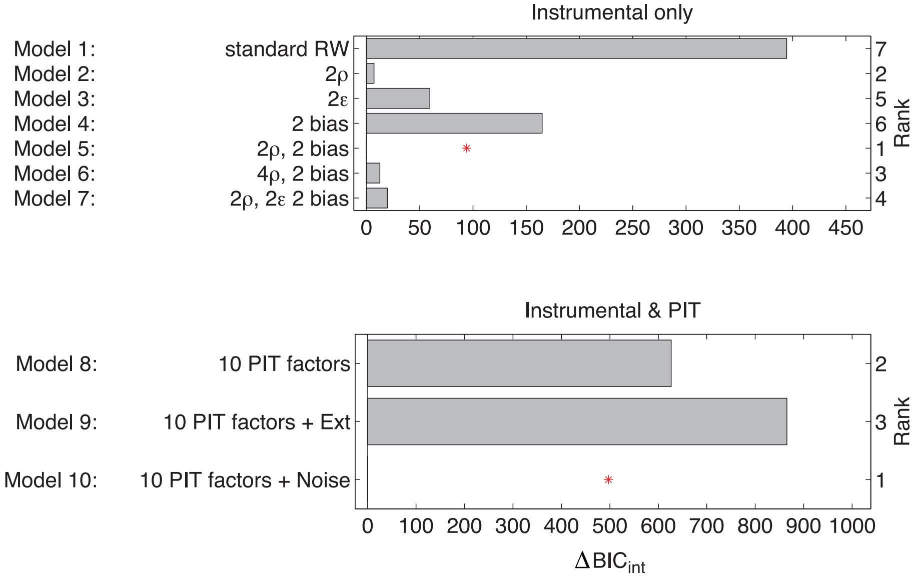

Figure 3. Model comparison. Each bar shows the differential $\mathrm{BIC}_{\mathrm{int}}$ score relative to the model with the lowest $\mathrm{BIC}_{\mathrm{int}}$ score (log e scale). Note that these $\mathrm{BIC}_{\mathrm{int}}$ scores are for the group as a whole. **Top**: Models 1–7 were fitted to the instrumental data only. Model 1 was a standard Rescorla-Wagner type model which forced rewards and punishments to be equally informative. It assumed equally fast learning about rewards and punishments, and no biases. Inclusion of either separate reward and punishment sensitivities (2ρ, Model 2) or separate biases in the approach and withdrawal blocks (Model 4) improved the fit. Separate learning rates for rewards and punishments (Model 3) did not improve the fit as much as separate reward and punishment sensitivities (Model 2). The best model (5) included a separate go bias in the approach and withdrawal blocks, and separate reward and punishment learning rates. Models that additionally allowed separate reinforcement sensitivities (Model 6), or separate learning rates (Model 7) in the approach and withdrawal blocks failed to improve the fit. **Bottom**: Comparison of models on both instrumental and PIT choice data jointly. Models 8–10 used the instrumental component of Model 5. Models 8–10 included ten Pavlovian factors, capturing the effect of each of the five Pavlovian stimuli in each of the two blocks. Model 9 allowed for extinction by including an exponential decay of the instrumental values during the PIT part of the task. Model 10 included random generalisation noise and provided the best fit.
doi:10.1371/journal.pcbi.1002028.g003

the task, and was fit to the choice data of all subjects. We used group-level Bayesian model comparison [66] to choose amongst a variety of model formulations (reporting $\Delta\mathrm{BIC}_{\mathrm{int}}$ scores relative to the final model), and ensured that inference yielded correct parameter estimates when run on surrogate data generated from the assumed underlying decision process.

Instrumental learning

The final model included 5 parameters associated directly with the instrumental requirements of the task. These comprise one learning rate ϵ; two parameters $\mathrm{bias}_{\mathrm{app}}$ and $\mathrm{bias}_{\mathrm{wth}}$ representing the bias towards go in the approach and withdrawal blocks; and two separate free parameters ρ_{rew} and ρ_{pun}, representing the effective strengths of rewards and punishments.

At a group level, subjects were biased against active withdrawal, but showed no bias for or against approach ($p = 8 \times 10^{-8}$ and $p = 0.70$ respectively, two-tailed t-test), the difference being significant ($p = 5 \times 10^{-5}$, ANOVA, Figure 4A). Withdrawal biases in the release and throw away experimental subgroups did not differ ($p = 0.62$, ANOVA), controlling for motor effects. The withdrawal bias accounts for the lower performance on go withdrawal in Figure 2A.

One concern is that differences in the biases might have masked differences in learning (i.e. the reward sensitivities) in the approach and withdrawal conditions. We tested this by allowing for separate reward and punishment sensitivities in the two conditions (Model 6) or separate learning rates (Model 7). The use of these extra parameters was structurally rejected by the model selection process ($\Delta\mathrm{BIC}_{\mathrm{int}} = 12.6; 19.7$ respectively for the purely instrumental trials); and the freedom to choose different parameter values in these conditions was duly not used (Figure 5). The absence of any difference in the *learning* parameters for approach and withdrawal suggests that the instrumental system treated approach and withdrawal entirely equally. We will see below that this was not true for the Pavlovian system.

Although, by design, rewards and punishments were equally informative, subjects chose to rely more on rewards than punishments (Figure 4B). Rewards had a stronger effect than punishments both at a group level and for all individual subjects, the difference being significant ($p < 1 \times 10^{-15}$, ANOVA). Indeed, the average punishment sensitivity was not distinguishable from zero ($p = 0.37$, two-tailed t-test). This remained true when we separately tested subjects who were given deterministic ($p = 0.34$, two-tailed t-test) and probabilistic ($p = 0.0627$, two-tailed t-test) feedback. Supplementary analyses (Text S1) excluded two further explanations for the punishment insensitivity: first, that it is due to choice perseverance (Figure S1 Text S1); and second that it is due to an emerging maximisation behaviour (Figure S2 in Text S1). Thus, it appears that the pattern seen in Figure 2B is indeed due to a differential sensitivity to rewards and punishments.

Generalization: Extinction versus noise

We next analysed the generalization of instrumental $\mathcal{Q}(s,a)$ values from the instrumental to the PIT blocks. Generalization could be imperfect in two ways - the starting $\mathcal{Q}(s,a)$ values in the PIT block could differ from the ending $\mathcal{Q}(s,a)$ values in the preceding instrumental block, and the $\mathcal{Q}(s,a)$ values could then decay over time or trials during the PIT block given the lack of information about the outcomes. We constructed models including such effects, and tested whether their excess complexity was outweighed by their fit to the data.

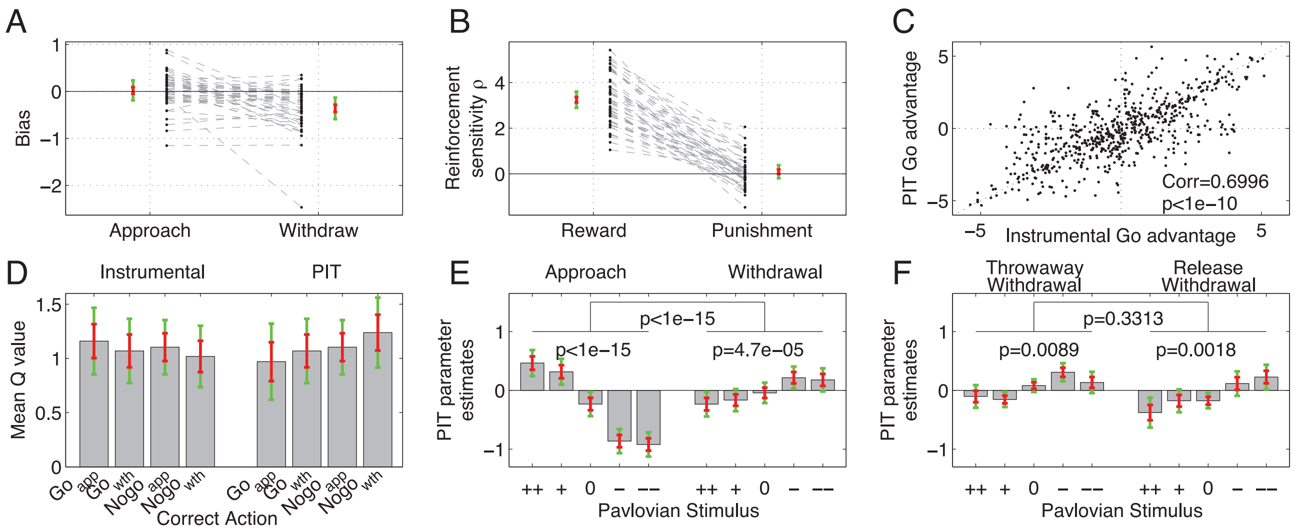

Figure 4. Instrumental model parameters. A: Go biases for the approach and withdrawal condition in the full experiment. Subjects were only biased against go, compared to nogo, in the withdrawal block. **B**: Reward and punishment sensitivity. Subjects were significantly more sensitive to rewards than punishments. **C**: Generalization noise. Effective $\mathcal{Q}$ value differences between go and nogo actions for all stimuli and subjects, at the end of instrumental learning and during the PIT block. Generalization seemed noisier when action preferences were weaker. **D**: Mean $\mathcal{Q}$ values of 'correct' (i.e. more frequently rewarded) actions. There was no difference, and all correct actions had *positive* expectations on average. **E**: PIT parameter estimates, correcting for instrumental learning, response biases and generalization noise. Positive Pavlovian stimuli enhanced approach go actions and inhibited withdrawal go, while negative Pavlovian stimuli inhibited approach go actions and enhanced withdrawal go actions. The interaction was highly significant, as were the two linear main effects. **F**: There was no difference between the effect of Pavlovian stimuli on throwaway versus release go actions (all p values in E and F are ANOVA). Throughout, grey bars are prior means with estimates of standard error (red) and 95% confidence interval (green). Black dots show individual data points, and individual subjects' parameters are connected by a dashed grey line in A and B.

doi:10.1371/journal.pcbi.1002028.g004

As expected from the stable raw probabilities of choosing the correct (i.e., more rewarded) option (Figure 2C), a model in which the instrumental $\mathcal{Q}(s,a)$ values decayed exponentially over time during the PIT block (mimicking extinction) did not provide a good account of the data (Model 9, compared to Model 10 $\Delta\mathrm{BIC}_{\mathrm{int}} = 865$).

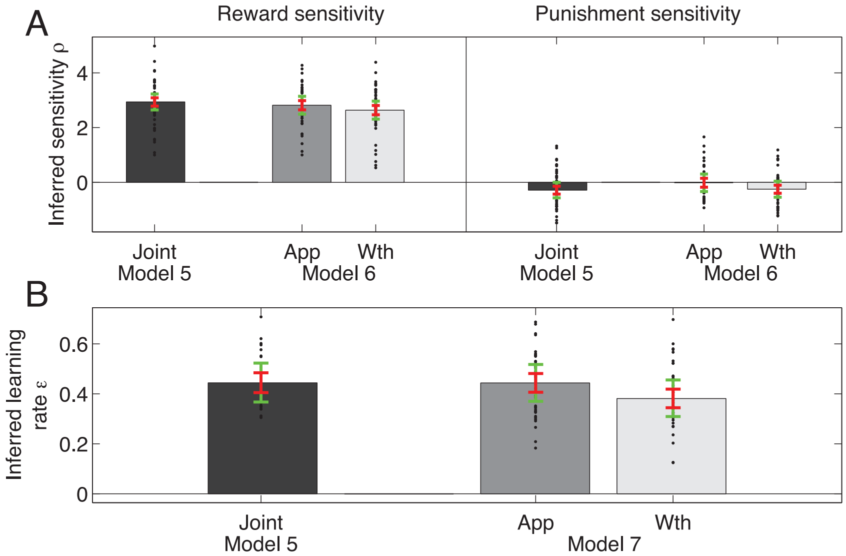

Figure 5. Reward sensitivities and learning rates in instrumental approach and withdrawal blocks do not differ. A: The dark bars show the reward (left) and punishment (right) sensitivities in Model 5, which collapses across approach and withdrawal conditions. The grey and light grey bars show the sensitivities when fit separately for approach and withdrawal blocks (Model 6). There is no difference between blocks; and the joint parameter differs from neither (all pairwise comparisons $p > .19$). **B**: Dark bar shows learning rate collapsed across both conditions in Model 5. Grey and light grey bars show learning rates when fit separately for approach and withdrawal condition. Again, no pairwise difference is significant (all $p > .2$). Throughout, black dots show individual data; bars show prior means and red and green error bars 1 estimated standard error and 95% confidence interval, respectively.

doi:10.1371/journal.pcbi.1002028.g005

Rather, the final model allowed for the addition of random generalization noise to each $\mathcal{Q}(s,a)$. These factors were drawn independently from the same normal distribution for all stimulus-action pairs, and the mean and variance of this distribution were both inferred without constraints (see Methods). Figure 4C visualizes the resulting changes; each dot represents the preference for the go action ($\mathcal{Q}(s,\text{go})-\mathcal{Q}(s,\text{nogo})$) for all subjects and all stimuli. The abscissa shows this at the end of the instrumental stage, the ordinate after addition of the noise for the PIT stage. Importantly, there was no systematic difference in mean correct action values either in the instrumental or PIT stage (Figure 4D).

Pavlovian-Instrumental transfer

We were mainly interested in the effect of the Pavlovian values on instrumental performance. We therefore fitted 10 unconstrained parameters to separately capture the influence of each of the five Pavlovian stimuli on instrumental go actions in both the approach and withdrawal condition.

All models accounted for performance in the PIT part by adding up instrumental and Pavlovian influences prior to taking a softmax [67,68]. This amounts to treating instrumental and the Pavlovian controllers as separate experts, each of which 'voted' for its preferred action. The model captured in detail, and thereby controlled for, variability in instrumental learning and generalization. The final model predicted the choices of every individual subject better than chance (binomial probability, $p<.0001$ for every subject, overall predictive probability 0.7544). The maximum a posteriori (MAP) estimates of this model's parameters painted a picture very similar to that seen in the raw data.

Figure 4E shows the parameters of the model related to the influence of each Pavlovian stimulus. The pattern mirrored that seen in the raw data: there are highly significant, and opposite, effects in the approach and withdrawal blocks, with appetitive stimuli (++ and +) promoting approach but inhibiting withdrawal; and aversive stimuli (-- and -) promoting withdrawal but inhibiting approach. At a single subject level, the effect in the approach block was seen in 45/46 subjects (98%), while it was seen in 30 subjects (65%) in the withdrawal block.

Since there was no difference in the learned value of go or nogo actions in either approach or withdrawal blocks, and in either the instrumental learning or the PIT stages (Figure 4D), any PIT effects are unlikely to be due to a preferential association of a Pavlovian stimulus with the learned value of an action. Rather, they reflect the approach or a withdrawal nature of the action.

We included two separate groups of subjects who either performed a throwaway withdrawal action, or a release withdrawal action. This was both to test the contribution of an approach/withdrawal component aimed at the Pavlovian stimuli tiling the background, and in recognition of the sophistication of defensive reactions [27]. Figure 4F shows that Pavlovian stimulus value had a significant, linear effect on both withdrawal action types, and that this overall linear effect did not differ between the two action types. At an individual level, linear correlations were positive for 16 (72%) and 14 (58%) subject in the release and throwaway condition, respectively.

Psychometric measures

No psychometric measure of anxiety or depression correlated with any of the parameters in the main model.

Discussion

Our task was designed to look inside the triads of valence, behavioural activation and inhibition, and specific actions associated with Pavlovian influences. This issue has been incompletely explored in the past. Either these triads as a whole have been investigated: aversive actions allowed avoidance of, or escape from, a negative reinforcer; appetitive actions, the acquisition of a reward [6,8,64], or, as in negative automaintenance [10], the relevant Pavlovian contingencies have been tightly embedded in the instrumental task. Here, we found that Pavlovian influences distinguished approach from withdrawal when carefully controlling for activation, for appetitive versus aversive instrumental motivation, and for details of the motor execution. Thus, for instance, a Pavlovian stimulus predicting reward had opposite effects on two different instrumental actions (approach and withdrawal) even though both those actions were themselves equally motivated by the acquisition of reward.

Approach and avoidance were defined in two parallel ways: by the cognitive label for the action ('throw away', 'collect') and by the relation to the stimulus (moving the mouse/finger towards or away from the stimulus). Our task did not set out to distinguish these two contributions (cognitive and motor), and we also did not attempt to quantify subjects' explicit insight into their strategies.

However, both possibilities are important. At a cognitive level, subjects should neglect the Pavlovian stimuli: by design, they are not informative about the instrumental task. Upon entering the PIT stage, subjects were also explicitly instructed to continue doing the instrumental task as before. If despite these facts subjects were cognitively swayed to include the irrelevant backgrounds in their goal-directed decision process, then our finding show that Pavlovian contingencies extend even into cognitive choices. This is of course consonant with a large number of behavioural irregularities in human decision making [12–15].

The motor aspects are equally interesting since they suggest a fine level of detail in the architecture of Pavlovian influences. There is quite some evidence for this; for instance, Pavlovian CRs are known to be highly adaptive to the details of the CS (for instance evoking a grooming conditioned response to a rat which functions as a food CS, rather than a gnawing CR [69]) and to the nature of the US [70]. In humans, a plexiglass positioned between subjects and an appetitive US abolishes an increased willingness to pay [71].

The performance on the purely instrumental portion of the task was also revealing. We observed a difference in the instrumental performance of approach and withdrawal action; and this came (unlike in previous tasks) after controlling for the motivational difference between approach and avoidance. Our model-based analysis revealed that the difference was not due to a difference in learning (i.e. a difference in the instrumental parameters relating reinforcements to performance), but due to a static bias against performing a withdrawal go action. Of course, like all other tasks, our instrumental task also had embedded Pavlovian contingencies, and, indeed, a Pavlovian suppression of active withdrawal by the overall appetitive framing of the task (subjects on average chose the correct, rewarded, action more often) could mirror what we saw in the PIT stage of the task. Alternatively, this could be the result of subjects' experiences upon entering an experimental situation in which they are given a computer mouse. We have interpreted such as bias in terms of evolutionary preparedness or programming [2,9,24,50,72]. That is, the flexibility of the arbitrary outcome-contingent mappings of instrumental control comes at the price of the experience necessary for it to be specified. Pavlovian priors substitute inflexible hard-wired choices that are immediately available for this flexible instrumental adaptativity with its potentially substantial sample complexity (i.e. the potential need for extended experience). Related biases are widely known: dogs will happily learn to run, but not to yawn, for food; teaching a rat to escape is easier than teaching it to avoid the shock [3,27,28]; humans perform active go responses slower if instructions are in terms of aversive feedback [51] or if they are followed by aversive information [73]. Finally, in humans, an instructed joystick approach response to a happy face is quicker than a withdrawal response, depending on the cognitive/affective label in a manner similar to our own findings here [74].

Alternative interpretations of the response bias include endowment effects [14], whereby an over-valuation of items notionally in one's possession makes one reluctant to give them up. This is unlikely because such a bias should be present across all instrumental stimuli, i.e. across both stimuli for which a go and a no-go is the more rewarded action (Figure 4). Another possibility is a frame dependence [13]—since we compared go with nogo rather than two alternative go actions against each other. The negative frame associated with sorting to remove bad mushrooms could have inhibited go actions.

Neurobiology

One of the central motivations for our investigation was the observation that the neural substrate does not respect the logical independence of reward/punishment and approach/withdrawal. Rather, as we have discussed, these are tied together, via the structure of the striatum and also specific neuromodulators.

While the neural basis for the promotion of approach responses by appetitive stimuli is known to involve both amygdala and striatum [62,63,75], the neural bases for the effects of aversive Pavlovian stimuli are less clear. There are no data on withdrawal responses per se, i.e. with positive expectations. Nevertheless, animal models, genetic studies and pharmacological manipulations suggest that serotonin plays a crucial role in the inhibition of active behaviours by aversive expectations [25,47,48,50,73,76–78]. In humans, there is evidence for the serotonergic mediation of the inhibition of active approach by aversive predictions [51], and of approach responses to stimuli that are predictive of negative reinforcement [73]. It should be noted, though, that, acting via the indirect path and D2 receptors, dopamine itself has also been suggested to be important in mediating 'nogo' behaviour due to punishments [18,53,79].

Aversive Pavlovian stimuli can also potentiate behaviour [1,64,80,81], with both serotonin and dopamine involved. Dopamine may have a dominant influence in this: it is both known to be released, and influential, in some aversive settings [82–85] and has a more evident relationship to vigour [33,34]. This observation has led to a re-interpretation of previous notions [43] of the opponency between dopamine and serotonin, putting an axis spanning invigoration and inhibition together with spanning reward and punishment [52].

Thus, the literature suggests three predictions for genetic correlates of the Pavlovian influences we observe. When considering these, the caveats concerning the interaction of genetic variation with psychopathology (e.g. anxiety or depression), and with development need to be kept in mind. Nevertheless, the conditioned suppression effect of aversive Pavlovian stimuli on approach should be enhanced by D2 receptors, and hence be positively related to D2 striatal receptor density thought to be modulated by C975T (rs6277; [17]). Second, conditioned suppression should be increased in subjects with higher serotonin levels, i.e. as might be the case with the less efficient (s) allelic variation of the serotonin reuptake transporter (5HTTLPR SLC6A4 [86]). Third, given dopamine's established positive correlation with approach and PIT [87,88], we expect genetic polymorphisms that boost DA levels, such as the SLC6A3 polymorphism of the dopamine transporter [89], to increase the impact of appetitive Pavlovian stimuli on approach. A similar effect may be expected from DARPP-32, although its closer relationship to synaptic plasticity would also suggest effects on instrumental learning [90–92].

Instrumental punishment insensitivity

Although the learning parameters associated with instrumental approach and withdrawal did not differ, the impact of rewards and punishments on the acquisition of responding was highly asymmetric. In general, subjects neglected punishments, whilst maintaining a fixed sensitivity to reward. This was gratuitous as, in our setting, rewards and punishments were equally informative. It is, however, the case that the optimal strategy can be arrived at by concentrating on either.

Subjects were not globally insensitive to punishments, as their choice behaviour in the Pavlovian learning was highly accurate both for rewards and punishments. Furthermore, it should be emphasized that ascribing punishments a value of zero outcome would still effectively behave as a punishment because a zero outcome is well below the average expectation of correct actions (Figure 4D) and as such would reduce the tendency to emit the action that caused it. The asymmetry has been noted before. Others have fitted models with separate learning rates for rewards and punishments and reported significantly slower learning rates for punishments than rewards [93,94]. In some restricted regimes, learning rates and inverse temperature parameters can trade off, and we explicitly tested both types of models to address this.

One potential confound is the emergence of determinism. Subject were instructed to perform choices relative to mushrooms. Real world mushrooms are either edible or poisonous, and this dichotomy may have predisposed subjects towards a deterministic, rather than a matching, strategy. (For instance, subjects may have chosen responses based on a classification of the mushrooms into 'good' and 'bad' ones, rather than on the particular value of a response for a mushroom.) Indeed, in RL settings it is typically optimal to start with a low, exploratory, sensitivity to outcomes, but to increase this over time to encourage exploitation, culminating in a deterministic strategy [2]. However, subjects did not behave deterministically at any point (Figure 2A) and supplementary analyses showed that the time-varying pattern of reinforcement sensitivities this would predict is not observed in the data (Text S1). A further potential confound is the average stay probability. If this were precisely half-way between the stay probabilities after rewards and punishments in Figure 2B, then rewards and punishments would have the same effect relative to the baseline, and hence arguably be equally informative. However, this argument would neglect the fact that the mean stay probability itself must be a function of the reinforcement history; and that this must be included in making inferences about the reinforcement sensitivity.

We have previously made the argument on theoretical grounds that part of the asymmetry observed in appetitive and aversive systems might be due to the inherent difference in how informative rewards and punishments are processed, enshrined again in the architecture of the striatum and neuromodulation [50]. Rewards tell us what to do; punishments tell us what not to do. The former is more informative in naturalistic settings where many options are available but only few are good. The fact that subjects gratuitously rely on rewards rather than on punishments in the present setting may reflect an implicit appreciation of this fact, although our findings are certainly in no way conclusive evidence. Interestingly, it is known that stronger optimality results can be shown for a stochastic learning automata rule called linear reward-inaction, which does not change propensities in the light of punishments but only rewards ([95,96]; also known as a benevolent automaton [97]), than for a rule that changes propensities for both.

Modelling

The computational model served several central roles. First, it encapsulated the manifold aspects of behaviour and learning *jointly*, thereby controlling for them: the bias against withdrawals is not a due to a difference in learning; and variations in learning or generalization do not account for the PIT effects we saw. Secondly, its close fit to the behaviour argues that the PIT effects can be accounted for by a simple superposition of an instrumental and a Pavlovian controller: the action propensities due to both controllers were simply multiplied (as additive factors in an exponential), rather than being allowed to interact in more complex ways.

The simplicity of this interaction eschews questions about peripheral versus central response competition, whether appetitive and aversive systems compete centrally [7], and whether Pavlovian learning is involved in instrumental learning [1]. It takes the view of multiple, separate controllers contributing in parallel [98], and weighting the ultimate choice by the reward expected from that choice. One alternative would be to weigh contributions by different controllers according to their certainty [99], although it is unclear how to compute the Pavlovian controller's certainty.

Limitations

There are various pressing directions for future studies. First, despite the role the architecture of decision-making has played in the argument, our work does not directly address the neural mechanisms concerned. These could be examined using imaging and pharmacological manipulations.

Second, our task was not designed to distinguish between outcome-specific and general mechanisms [63,75] as we relied on one, monetary, outcome throughout. Studying different outcomes is important, given evidence for partly parallel pathways through different nuclei of the amygdala and different targets in the nucleus accumbens [100,101].

Third, we are missing one crucial further orthogonalization to do with the overall framing of the instrumental task. It is important to consider the case in which subjects can at best avoid losing money by doing the correct action [51]. We would expect punishment to maintain its instrumental force in this case; but there could also be a systematic difference in the nature of the Pavlovian influences.

Conclusion

Pavlovian responses are believed to be hard-wired to reflect evolutionarily appropriate attitudes to predictions, being highly adaptive and sensitive to environmental structures [102]. Here, we showed that Pavlovian influences on instrumental behaviour depend on the intrinsic affective label of an action, independent of its learned reward expectation.

It has long been known that prepared or compatible [27,69] behaviours are easier targets for instrumental conditioning. These intrinsic biases, or priors, may serve a crucial function both by reducing the need for collecting data (i.e. sample complexity) about the effects of actions, and by reducing the need for executing complex processing necessary to work out optimal actions (i.e. computational complexity). Both of these can be expensive or dangerous, particularly in an aversive context. Our findings sharpen the understanding of the relative contribution of Pavlovian and instrumental contingencies in general tasks. We showed clearly that the interaction of Pavlovian and instrumental behaviours is organized along the lines of appetitive and aversive motivational systems, and that a critical contributor to this is the affective nature of actions.

Methods

Subjects and procedure

54 healthy subjects of central European origin were recruited from the Berlin area. Subjects were screened for a personal history of neurological, endocrine, cardiac and psychiatric disorders (SCID-I screening questionnaire), and for use of drugs and psychotropic medication in the past 6 months. Subjects received performance-dependent compensation (5–32 Euro) for participation. Three subjects did not meet inclusion criteria and one subject did not complete the task; the data for three further subjects were lost due to a programming error. One further subject was excluded from the analysis because the instrumental task was not satisfactorily performed. The 46 remaining subjects were 25.3 ± 4.7 years old. 59% were female ($n=27$). The study was approved by the local Ethics Committee and was in accord with the Declaration of Helsinki 2008. Subjects were given detailed information and gave written consent. They were seated comfortably at a table in front of a laptop with headphones and used a mouse with their dominant hand to indicate their choices. The amount earned was indicated by the computer, and the sum paid in cash at the end of the session. The computer task was followed by completion of self-rating scales.

Task description

The task was written using Matlab and Psychtoolbox (http://psychtoolbox.org). It consisted of one approach and one withdrawal block separated by a 2 minute break. Each block was in turn divided into a instrumental training, a Pavlovian training and a PIT part. Table 1 illustrates this.

Instrumental training. The instrumental task was framed in terms of a mushroom collecting and sorting task. Instrumental stimuli were generic, coloured mushroom shapes. Trials started when subjects clicked in a central square (Figure 1A). In the approach block, instrumental stimuli $s^{\mathcal{I}}_{1,2,3}$ and $s^{\mathcal{I}}_{4,5,6}$ (with subscripts indicating the *identity* of stimuli, not the time of presentation) were then presented to one side, surrounded by a blue frame (Figure 1A, middle column, top). Subjects indicated that they wanted to collect the mushroom by moving the cursor onto the mushroom and clicking on it (approach go). They could also decide not to collect the mushroom by doing nothing for 1.5 seconds (approach nogo). At the end of each trial (after a click for go trials or after 1.5 s for nogo trials respectively), the stimulus disappeared and the outcome was shown in the middle of the screen (Figure 1A). In the withdrawal blocks, instrumental stimuli $s^{\mathcal{I}}_{7,8,9}$ and $s^{\mathcal{I}}_{10,11,12}$ were presented. Subjects chose whether to throw away mushrooms (withdrawal go) or do nothing (withdrawal nogo). Two different withdrawal go actions were tested. The 'throwaway' group ($n=24$) had to click in a blue frame located on the opposite side of the stimulus (see Figure 1A, middle column, middle). The 'release' ($n=22$) group was instructed to press *and hold* the mouse button after clicking in the central square to begin the trial. The mushroom was then presented underneath the cursor (Figure 1A, middle column, bottom), and they could throw away a mushroom by releasing the button (withdrawal go) or not throw away the mushroom by not releasing (withdrawal nogo) until 1.5 seconds had elapsed. Each block contained three "good" ($s^{\mathcal{I}}_{1,2,3}$ and $s^{\mathcal{I}}_{7,8,9}$) and three "bad" ($s^{\mathcal{I}}_{4,5,6}$ and $s^{\mathcal{I}}_{10,11,12}$) mushrooms, randomly selected from the pool of 12 stimuli. Subjects were given explicit reinforcing feedback after every choice ('Correct, +20 cents' or 'Wrong. −20 cents'), either deterministically ($n=19$) or probabilistically ($n=27$), but were not told which mushrooms were good or bad. Correct trials were those on which subjects threw away a bad or kept a good mushroom, and those on which they collected a good or refrained from collecting a bad mushroom. Importantly, this means that correct go actions of both types (approach ('collect') and withdraw ('throw away')) were followed by both rewards and punishments. Thus, the reinforcement expectancies of correct approach and withdrawal actions were equal and positive on average. Similarly, incorrect actions of both types were also followed by rewards and punishments, but more by the latter than the former. To ensure replicability across experimental designs, four experimental configurations were included, crossing deterministic/probabilistic instrumental feedback and the two withdrawal action types ('throw away' or 'release'). These manipulations are beyond the mathematical model described below, and thus should not affect our findings. We present both data for all subjects and, testing internal consistency, across the four groups. 10 subjects were in the deterministic throwaway group, 9 in the deterministic release, 14 in the probabilistic throwaway and 13 in the probabilistic release group. One-way ANOVA comparisons of MAP parameter estimates from the most parsimonious model (Model 10; see below) for deterministic and probabilistic feedback did not reveal any significant differences.

Pavlovian training. Five compound Pavlovian stimuli consisting of a fractal visual stimulus (Figure 1B) and a tone were classically conditioned. Each stimulus was presented 20 times and deterministically followed, 1 second later, by the associated outcome. Outcome presentation lasted 1.5 seconds. Outcomes for the best ($s^{\mathcal{P}}_{++}$), good ($s^{\mathcal{P}}_{+}$), neutral ($s^{\mathcal{P}}_{0}$), bad ($s^{\mathcal{P}}_{-}$) and worst ($s^{\mathcal{P}}_{--}$) stimuli were, respectively, gains of 100 cents, 10 cents, zero, and losses of 10 and 100 cents. To ensure that subjects paid attention, every fifth trial was a query trial in which subjects had to choose between two Pavlovian stimuli (Figure 1C). No feedback was given in these trials, but subjects were instructed that the choices would contribute to their compensation.

Pavlovian-Instrumental transfer. In the final part of each block, the instrumental task was presented in extinction and on the background of Pavlovian stimuli (Figure 1D). Subjects were instructed to continue doing the instrumental task; that choices were still earning them the same outcomes and were being counted, but that they would not be told about the outcomes. Note, importantly, that the Pavlovian stimulus was presented over the entire background, and as such could not by itself modulate the directionality of actions.

Psychometric measurements. After completing the tasks, subjects completed self-rating scales (Beck Depression Inventory II (BDI), Beck Anxiety Inventory (BAI), State-Trait Anxiety Inventory STAI [103–105]), followed by the administration of clinician rated scales (Montgomery-Ashberg Depression Rating Scale (MADRS), Hamilton Depression Scale (HamD), Structured Interview for the Hamilton Anxiety Scale (SIGHA) and Clinical Global Impression (CGI) [106–108]).

Models

We modified a standard reinforcement learning model to capture the behavioural choices in the experiment. We first describe the main model, and then the alternative control models. Considering first the instrumental part, let $s^{\mathcal{I}}_t$ be the instrumental stimulus (out of up to 12; i.e. the subscript t now designates *time* rather than identity as in Table 1) presented at trial t, and a_t the action (choice) on that trial. An action can be one of four types: go withdrawal and nogo withdrawal in the withdrawal block, and go approach and nogo approach in the approach block. Let also $r_t \in \{-1,1\}$ be the reinforcement obtained, either -1 for a punishment, or $+1$ for a reward. We write the probability of action a_t in the presence of stimulus $s^{\mathcal{I}}_t$ as a standard probabilistic function of i) the reinforcement expectations $\mathcal{Q}_t(s^{\mathcal{I}}_t,a_t)$ associated with that pair on that trial, and ii) a time-invariant, fixed, response bias $b(a_t)$:

$$\mathcal{W}^{\mathcal{I}}(s^{\mathcal{I}}_t,a_t)=\mathcal{Q}_t(s^{\mathcal{I}}_t,a_t)+b(a_t) \tag{1}$$

$$p(a_t|s^{\mathcal{I}}_t)=\frac{\exp\left(\mathcal{W}^{\mathcal{I}}(s^{\mathcal{I}}_t,a_t)\right)}{\sum_{a'}\exp\left(\mathcal{W}^{\mathcal{I}}(s^{\mathcal{I}}_t,a')\right)} \tag{2}$$

where $\mathcal{W}^{\mathcal{I}}$ is the instrumental weight of action a_t, and where the variable $b(a_t)$ can take on value bias_{wth} for withdrawal go actions, or bias_{app} for the approach go actions. It is always zero for the nogo action. There was no delayed outcome in the instrumental task, and the expectations were thus constructed by a Rescorla-Wagner-like rule with a fixed learning rate ϵ. The immediate, intrinsic, value of the reinforcements delivered in the experiment may have different meaning for different subjects. To measure this effect, we added two further parameters: the reward sensitivity ρ_{rew} and the punishment sensitivity ρ_{pun}, yielding an update equation for the expectations:

$$\mathcal{Q}_{t+1}(s^{\mathcal{I}}_t,a_t)=\mathcal{Q}_t(s^{\mathcal{I}}_t,a_t)+\epsilon\left(R_t-\mathcal{Q}_t(s^{\mathcal{I}}_t,a_t)\right)$$

$$R_t=\begin{cases}\rho_{\text{rew}} & \text{if } r_t>0\\ \rho_{\text{pun}} & \text{if } r_t<0\end{cases}$$

This is model 5 in Table 2, which has the lowest BIC_{int} score (see below). Alternative models tested on the instrumental data only are as follows: Model 1 assumes that $-\rho_{\text{pun}}=\rho_{\text{rew}}=\beta$, and that $\text{bias}_{\text{wth}}=\text{bias}_{\text{app}}=0$. Model 2 allows only for separate reward and punishment sensitivities and model 4 for separate biases. Model 3 again assumes $-\rho_{\text{pun}}=\rho_{\text{rew}}=\beta$, and that $\text{bias}_{\text{wth}}=\text{bias}_{\text{app}}=0$, but allows for two separate learning rates, i.e. ϵ in Equation 3 is replaced by ϵ_{rew} on trials where $r_t=1$, and by ϵ_{pun} on trials where $r_t=-1$. Model 6 and 7 are expansions of the final model, allowing for separate reward and punishment sensitivities (model 6) and for separate learning rates (model 7) in the approach and withdrawal conditions.

Our main measure of interest is the effect of Pavlovian stimuli on the approach and withdrawal actions. Let additionally $s^{\mathcal{P}}_t$ be the Pavlovian stimulus on trial t. We can then write an equation similar to equation 2 for the trials where both instrumental and Pavlovian stimuli were present, but including a term $f(a,s^{\mathcal{P}}_t)$ that quantifies the effect of the particular Pavlovian stimulus $s^{\mathcal{P}}_t$ on the action a. This means that the action weights due to the instrumental and Pavlovian controllers are added inside the exponent of equation 2, and that thus the probabilities each controller attaches to a particular action are multiplied and renormalized. The two controllers are therefore treated as two distinct entities, each separately voting for a particular action to be emitted. The influence of each system on action choice is *relative* to the strength with which the other enhances one particular action. We write the PIT weight of action a as:

$$\mathcal{W}^{\text{PIT}}(a,s^{\mathcal{I}},s^{\mathcal{P}})=\mathcal{W}^{\mathcal{I}}(s^{\mathcal{I}},a)+f(a,s^{\mathcal{P}}) \tag{3}$$

Here we force $f(\text{nogo},s^{\mathcal{P}})=0$ at all times. The go values $f(\text{go},s^{\mathcal{P}})$ can take on 10 separate, inferred, values, meaning that there is one separate parameter for each of the five Pavlovian stimuli $s^{\mathcal{P}}$ in each of the two blocks. Each of these parameters captures how much $s^{\mathcal{P}}$ boosts the go over the nogo action (if $f(\text{go},s^{\mathcal{P}})>0$) or the inverse (if $f(\text{go},s^{\mathcal{P}})<0$). Note that because these are separately inferred, independent, parameters, this formulation does not impose any assumptions about the effect of the value of the stimulus $s^{\mathcal{P}}$, or about the relative effect of different stimuli $s^{\mathcal{P}}$ with different values. Hence, this controls for variation in learning during the Pavlovian training block (though the query trials indicate that learning was very robust).

Equation 3 (Model 8 in Table 2) assumes that the stimulus-action values $\mathcal{Q}(s^{\mathcal{I}},a)$ at the end of the instrumental block are perfectly and exactly generalized to the PIT block. We first tested an alternative model (Model 9 in Table 2) that included an exponential extinction

Table 2. Parameters contained in each of the models in Figure 3.

Model	Data			Parameters		Generalization	$\mathbf{BIC_{int}}$
1	instrumental	ϵ	β				5000
2	instrumental	ϵ	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$				4613
3	instrumental	$\epsilon_{\mathrm{rew}},\epsilon_{\mathrm{pun}}$	β				4665
4	instrumental	ϵ	β	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$			4771
5	instrumental	ϵ	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$			4606
6	instrumental	ϵ	$\rho^{\mathrm{app}}_{\mathrm{rew}},\rho^{\mathrm{app}}_{\mathrm{pun}},\rho^{\mathrm{wth}}_{\mathrm{rew}},\rho^{\mathrm{wth}}_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$			4618
7	instrumental	$\epsilon_{\mathrm{app}},\epsilon_{\mathrm{wth}}$	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$			4626
8	instr&PIT	ϵ	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$	10 separate $f(\mathrm{go},s^{\mathcal{P}})$	exact	17396
9	instr&PIT	ϵ	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$	10 separate $f(\mathrm{go},s^{\mathcal{P}})$	extinction	17634
10	instr&PIT	ϵ	$\rho_{\mathrm{rew}},\rho_{\mathrm{pun}}$	$\mathrm{bias}_{\mathrm{app}},\mathrm{bias}_{\mathrm{wth}}$	10 separate $f(\mathrm{go},s^{\mathcal{P}})$	noisy	16769

doi:10.1371/journal.pcbi.1002028.t002

factor, letting the $\mathcal{Q}$ values decay on each PIT trial by $\mathcal{Q}_{t+1}(a_t,s_t^{\mathcal{I}})=\alpha\mathcal{Q}_t(a_t,s_t^{\mathcal{I}})$ with $0\le\alpha\le 1$. Next, we tested the model described in the main text (Model 10 in Table 2), which allowed for a fixed, Gaussian random offset between the effective $\mathcal{Q}$ values in the instrumental and PIT stages, i.e. we wrote:

$$\mathcal{W}^{\mathrm{PIT}}(a_t,s_t^{\mathcal{I}},s_t^{\mathcal{P}})=\mathcal{W}_t^{\mathcal{I}}(s_t^{\mathcal{I}},a_t)+f(a_t,s_t^{\mathcal{P}})+\eta(a_t,s_t^{\mathcal{I}})$$

The noise factor $\eta(a,s_t^{\mathcal{I}})$ took on value 0 for the nogo action (akin to the bias and f variables). It took on a separate value—which was inferred as a separate parameter—for each subject and each stimulus. However, all stimuli shared the same prior distribution for this noise variable. That is, in the E step of our EM procedure, we fitted one Gaussian mean and variance to all the η's that had been inferred for all stimuli for all subjects. In this sense, the generalization factors η were drawn from one Gaussian prior whose mean and variance were fitted just like the mean and variance of the other parameters.

Model fitting procedure

For each subject, each model specifies a vector of parameters $\mathbf{h}$. Assuming Gaussian prior distributions $p(\mathbf{h}|\boldsymbol{\theta})$, we find the maximum a posteriori estimate $\mathbf{m}_i$ of the parameters for each subject i:

$$\mathbf{m}_i=\underset{\mathbf{h}}{\operatorname{argmax}}\,p(\mathbf{A}_i|\mathbf{h})p(\mathbf{h}|\boldsymbol{\theta})$$

where $\mathbf{A}_i$ are all actions by the i^{th} subject. We assume that actions are independent (given the stimuli, which we omit for notational clarity), and thus factorize over trials. The prior distribution on the parameters mainly serves to regularise the inference and prevent parameters that are not well-constrained from taking on extreme values. We set the parameters of the prior distribution $\boldsymbol{\theta}$ to the maximum likelihood given all the data by *all* the N subjects:

$$\hat{\boldsymbol{\theta}}^{ML}=\underset{\boldsymbol{\theta}}{\operatorname{argmax}}\,p(\mathcal{A}|\boldsymbol{\theta})$$
$$=\underset{\boldsymbol{\theta}}{\operatorname{argmax}}\left(\prod_{i=1}^{N}\int d^N\mathbf{h}_i p(\mathbf{A}_i|\mathbf{h}_i)p(\mathbf{h}_i|\boldsymbol{\theta})\right)$$

where $\mathcal{A}=\{\mathbf{A}_i\}_{i=1}^{N}$. This maximisation is straightforwardly achieved by Expectation-Maximisation [109]. We use a Laplacian approximation for the E-step at the k^{th} iteration:

$$p(\mathbf{h}|\mathbf{A}_i)\approx\mathcal{N}(\mathbf{m}_i^{(k)},\Sigma_i^{(k)})$$
$$\mathbf{m}_i^{(k)}=\underset{\mathbf{h}}{\operatorname{argmax}}\,p(\mathbf{A}_i|\mathbf{h})p(\mathbf{h}|\boldsymbol{\theta}^{(k-1)})$$

where $\mathcal{N}(\cdot)$ denotes a normal distribution over $\mathbf{h}$ with mean $\mathbf{m}_i^{(k)}$ and $\Sigma_i^{(k)}$ is the second moment around $\mathbf{m}_i^{(k)}$, which approximates the variance, and thus the inverse of the certainty with which the parameter can be estimated. Finally, the hyperparameters $\boldsymbol{\theta}$ are estimated by setting the mean $\boldsymbol{\mu}$ and the (factorized) variance $\boldsymbol{\nu}^2$ of the prior distribution to:

$$\boldsymbol{\mu}^{(k)}=\frac{1}{N}\sum_i\mathbf{m}_i^{(k)}$$
$$(\boldsymbol{\nu}^{(k)})^2=\frac{1}{N}\sum_i\left[(\mathbf{m}_i^{(k)})^2+\Sigma_i^{(k)}\right]-(\boldsymbol{\mu}^{(k)})^2$$

ansformed before inference to enforce constraints. Unconstrained parameters are inferred in their native space. These model fitting procedures were verified on surrogate data generated from a known decision process.

Model comparison

We fitted a large number of different models to the data, and some of these models differ in their flexibility. For instance, Model 8, which assumes that the instrumental $\mathcal{Q}$ values are generalized exactly to the PIT stage is much less flexible than models 9–10, which allow for an offset. It is important to choose that model which is flexible enough to explain the data, but not so flexible that it would also fit very different data equally well [109].

Ideally, this is achieved by computing the posterior log likelihood $\log p(\mathcal{M}|\mathcal{A})$ of each model $\mathcal{M}$ given all the data $\mathcal{A}$. As we have no prior on the models themselves (testing only models we believe are equally likely a priori), we instead examine the model log likelihood $\log p(\mathcal{A}|\mathcal{M})$ directly. This quantity can be approximated in two steps. First, the integral over $\boldsymbol{\theta}$ [110]:

$$\log p(\mathcal{A}|\mathcal{M})=\int d\boldsymbol{\theta}p(\mathcal{A}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{M})$$
$$\approx-\frac{1}{2}\mathrm{BIC}_{\mathrm{int}}=\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML})-\frac{1}{2}|\mathcal{M}|\log(|\mathcal{A}|)$$

Importantly, however, $\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML})$ is not the sum of individual likelihoods, but in turn an integral over the parameters of each individual subject:

$$\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}) = \sum_i \log \int d\mathbf{h} p(\mathbf{A}_i|,\mathbf{h}) p(\mathbf{h}|\hat{\boldsymbol{\theta}}^{ML})$$

$$\approx \sum_i \log \frac{1}{K} \sum_{k=1}^{K} p(\mathbf{A}_i|\mathbf{h}^k)$$

The second line shows that we approximated the integrals by (importance) sampling K times from the empirical prior distribution $\mathbf{h}^k \sim p(\mathbf{h}|\hat{\boldsymbol{\theta}}^{ML})$ [109]. These samples were then also used to derive the error bars as the second moments around the maximum:

$$\frac{\partial^2 p(\mathcal{A}|\boldsymbol{\theta})}{\partial h_l \partial h_m}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{ML}} \approx \frac{1}{\delta^2}\Big[p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}+\delta\mathbf{e}_l) -$$

$$2p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}) + p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}-\delta\mathbf{e}_m)\Big]$$

where $\mathbf{e}_l$ is a vector of zeros of the same dimension as $\boldsymbol{\theta}$ with only entry l set to one. The shifted likelihoods can be easily computed by re-weighting the K samples drawn before:

$$\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}+\delta\mathbf{e}_l) \approx \sum_i \log \sum_{k=1}^{K} p(\mathbf{A}_i|\mathbf{h}^k) w_{ik}^l$$

$$\tilde{w}_{ik}^l = \frac{p(\mathbf{h}^k|\hat{\boldsymbol{\theta}}^{ML}+\delta\mathbf{e}_l)}{p(\mathbf{h}^k|\hat{\boldsymbol{\theta}}^{ML})}$$

$$w_{ik}^l = \frac{\tilde{w}_{ik}^l}{\sum_{k'} \tilde{w}_{ik'}^l}$$

Note that while this model comparison procedure does give a good *comparative* measure of model fit, we still need an absolute measure to ensure that the best model does indeed provide a model fit that is adequate (even the best might be bad). Given each subject's MAP parameter estimate, we compute the total "predictive probability":

$$p(\mathcal{A}|\{\mathbf{h}_i\}_{i=1}^N) = \prod_{i=1}^{N} \prod_{t=1}^{T} p(a_t^i|s_t^{\mathcal{I}},\mathbf{h}_i) \qquad (4)$$

where we suppressed the dependence on stimuli on the LHS for clarity. We note that $p(a_t|s_t^{\mathcal{I}})$ depends on the parameters $\mathbf{h}_i$, which have been fitted to the data. We term it a predictive probability in the sense that it predicts a subject's choice at time t given that subject's *past* behaviour. We emphasize however, that this does depend on the MAP parameters $\mathbf{h}^i$ fitted to that subjects' entire choice dataset. Finally, we test whether the expected number of choices predicted correctly exceeds that expected by chance (using a binomial test). The overall predictive probability is given by the geometric mean over all choices and subjects: $\sqrt[TN]{p(\mathcal{A}|\{\mathbf{h}_i\}_{i=1}^N)}$.

Supporting Information

Text S1 Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding.
(PDF)

Acknowledgments

We would like to thank Bernard Balleine, Molly Crockett, Anthony Dickinson, Marc Guitart-Masip, Jon Roiser and Ben Seymour for helpful discussions and critically commenting on the manuscript.

Author Contributions

Conceived and designed the experiments: QJMH RC MG EF AH RJD PD. Performed the experiments: MG EF. Analyzed the data: QJMH RC PD. Contributed reagents/materials/analysis tools: QJMH. Wrote the paper: QJMH RC MG EF AH RJD PD.

References

1. Rescorla RA, Solomon RL (1967) Two-process learning theory: Relationships between pavlovian conditioning and instrumental learning. Psychol Rev 74: 151–182.
2. Sutton RS, Barto AG (1998) Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press, Available: http://www.cs.ualberta.ca/sutton/book/the-book.html.
3. Bouton ME (2006) Learning and Behavior: A Contemporary Synthesis. Sinauer.
4. Dickinson A (1980) Contemporary animal learning theory. Cambridge, UK: Cambridge University Press.
5. Gray JA (1991) The psychology of fear and stress, volume 5 of *Problems in the behavioural sciences*. Cambridge, UK: Cambridge University Press, 2 edition.
6. Estes W, Skinner B (1941) Some quantitative aspects of anxiety. J Exp Psychol 29: 390–400.
7. Dickinson A, Dearing MF (1979) Appetitive-aversive interactions and inhibitory processes. In: Dickinson A, Boakes RA, eds. Mechanisms of learning and motivation. Hillsdale, NJ: Erlbaum. pp 203–231.
8. Lovibond PF (1983) Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. J Exp Psychol Anim Behav Process 9: 225–247.
9. Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. Neural Netw 19: 1153–1160.
10. Williams DR, Williams H (1969) Auto-maintenance in the pigeon: sustained pecking despite contingent non-reinforcement. J Exp Anal Behav 12: 511–520.
11. Hershberger WA (1986) An approach through the looking-glass. Anim Learn Behav 14: 443–51.
12. Ainslie G (2001) Breakdown of will. Cambridge University Press.
13. Martino BD, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. Science 313: 684–687.
14. Kahneman D, Knetsch J, Thaler R (1990) Experimental tests of the endowment effect and the Coase theorem. J Polit Econ 98: 1325.
15. Ariely D (2008) Predictably irrational: The hidden forces that shape our decisions. HarperCollins London.
16. Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in Parkinsonism. Science 306: 1940–3.
17. Frank MJ, Hutchison K (2009) Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. Neuroscience 164: 131–140.
18. Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci 17: 51–72.
19. Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27: 7105–7116.
20. Drevets WC, Gautier C, Price JC, Kupfer DJ, Kinahan PE, et al. (2001) Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. Biol Psychiatry 49: 81–96.
21. Heinz A (2002) Dopaminergic dysfunction in alcoholism and schizophrenia–psychopathological and behavioral correlates. Eur Psychiatry 17: 9–16.
22. Boileau I, Assaad JM, Pihl RO, Benkelfat C, Leyton M, et al. (2003) Alcohol promotes dopamine release in the human nucleus accumbens. Synapse 49: 226–231.
23. Robinson TE, Berridge KC (2003) Addiction. Annu Rev Psychol 54: 25–53.
24. Huys QJM (2007) Reinforcers and control. Towards a computational ætiology of depression. Ph.D. thesis, Gatsby Computational Neuroscience Unit, UCL, University of London. Available: http://www.gatsby.ucl.ac.uk/qhuys/pub.html.
25. Dayan P, Huys QJM (2008) Serotonin, inhibition, and negative mood. PLoS Comput Biol 4: e4.
26. Cox SML, Benkelfat C, Dagher A, Delaney JS, Durand F, et al. (2009) Striatal dopamine responses to intranasal cocaine self-administration in humans. Biol Psychiatry 65: 846–850.
27. Bolles RC (1970) Species-specific defense reactions and avoidance learning. Psychol Rev 77: 32–48.
28. Seligman ME (1970) On the generality of the laws of learning. Psychol Rev 77: 406–18.

29. Wise R (1982) Neuroleptics and operant behavior: The anhedonia hypothesis. Behav Brain Sci 5: 39–87.
30. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38: 329–337.
31. Knutson B, Greer SM (2008) Anticipatory affect: neural correlates and consequences for choice. Philos Trans R Soc Lond B Biol Sci 363: 3771–3786.
32. Carter CJ, Pycock CJ (1978) Differential effects of central serotonin manipulation on hyperactive and stereotyped behaviour. Life Sci 23: 953–960.
33. Murschall A, Hauber W (2006) Inactivation of the ventral tegmental area abolished the general excitatory influence of pavlovian cues on instrumental performance. Learn Mem 13: 123–126.
34. Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191: 507–520.
35. 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 Brain Res Rev 31: 6–41.
36. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. J Neurosci 16: 1936–47.
37. Suri RE, Schultz W (1999) A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 91: 871–890.
38. Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412: 43–48.
39. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304: 452–4.
40. Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47: 129–141.
41. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9: 1057–1063.
42. Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10: 1615–1624.
43. Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15: 603–16.
44. Takase LF, Nogueira MI, Baratta M, Bland ST, Watkins LR, et al. (2004) Inescapable shock activates serotonergic neurons in all raphe nuclei of rat. Behav Brain Res 153: 233–239.
45. Takase LF, Nogueira MI, Bland ST, Baratta M, Watkins LR, et al. (2005) Effect of number of tailshocks on learned helplessness and activation of serotonergic and noradrenergic neurons in the rat. Behav Brain Res 162: 299–306.
46. Bolles R, Riley A (1973) Freezing as an avoidance response: Another look at the operant-respondent distinction* 1. Learn Motiv 4: 268–275.
47. Soubrié P (1986) Reconciling the role of central serotonin neurons in human and animal behaviour. Behav Brain Sci 9: 319–364.
48. Deakin JFW, Graeff FG (1991) 5-HT and mechanisms of defence. J Psychopharmacol 5: 305–16.
49. Cools R, Robinson OJ, Sahakian B (2008) Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. Neuropsychopharmacology 33: 2291–2299.
50. Dayan P, Huys QJM (2009) Serotonin in affective control. Annu Rev Neurosci 32: 95–126.
51. Crockett MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. J Neurosci 29: 11993–11999.
52. Boureau YL, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36: 74–97.
53. Hikida T, Kimura K, Wada N, Funabiki K, Nakanishi S (2010) Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. Neuron 66: 896–907.
54. Lobo MK, Covington HE, Chaudhury D, Friedman AK, Sun H, et al. (2010) Cell type-specific loss of bdnf signaling mimics optogenetic control of cocaine reward. Science 330: 385–390.
55. Alexander G, Crutcher M (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13: 266–271.
56. Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35: 4–26.
57. Faure A, Reynolds SM, Richard JM, Berridge KC (2008) Mesolimbic dopamine in desire and dread: enabling motivation to be generated by localized glutamate disruptions in nucleus accumbens. J Neurosci 28: 7184–7192.
58. Faure A, Richard JM, Berridge KC (2010) Desire and dread from the nucleus accumbens: cortical glutamate and subcortical gaba differentially generate motivation and hedonic impact in the rat. PLoS One 5: e11223.
59. Bandler R, Shipley MT (1994) Columnar organization in the midbrain periaqueductal gray: modules for emotional expression? Trends Neurosci 17: 379–389.
60. Gray JA, McNaughton N (2003) The neuropsychology of anxiety. Oxford University Press, 2nd edition.
61. Dickinson A, Pearce J (1977) Inhibitory interactions between appetitive and aversive stimuli. Psychol Bull 84: 690–711.
62. Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian-instrumental transfer. J Neurosci 28: 360–368.
63. Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP (2008) The neural mechanisms underlying the influence of pavlovian cues on human decision making. J Neurosci 28: 5861–5866.
64. Overmier J, Bull J, Pack K (1971) On instrumental response interaction as explaining the influences of Pavlovian Cs+ s upon avoidance behavior. Learn Motiv 2: 103–112.
65. Hirsch S, Bolles R (1980) On the ability of prey to recognize predators. Z Tierpsychol 54: 71–84.
66. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46: 1004–1017.
67. Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-Nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84: 581–617.
68. Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84: 555–579.
69. Timberlake W, Grant D (1975) Auto-shaping in rats to the presentation of another rat predicting food. Science 190: 690–692.
70. Holland PC (1977) Conditioned stimulus as a determinant of the form of the Pavlovian conditioned response. J Exp Psychol Anim Behav Process 3: 77–104.
71. Bushong B, King L, Camerer C, Rangel A (2010) Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay. Am Econ Rev 100: 1556.
72. Dickinson A, Smith J, Mirenowicz J (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 114: 468–483.
73. Cools R, Roberts AC, Robbins TW (2008) Serotoninergic regulation of emotional and behavioural control processes. Trends Cogn Sci 12: 31–40.
74. Roelofs K, Minelli A, Mars RB, van Peer J, Toni I (2009) On the neural control of social emotional behavior. Soc Cogn Affect Neurosci 4: 50–58.
75. Corbit LH, Balleine BW (2005) Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. J Neurosci 25: 962–970.
76. Tye NC, Everitt BJ, Iversen SD (1977) 5-hydroxytryptamine and punishment. Nature 268: 741–743.
77. Lister S, Pearce JM, Butcher SP, Collard KJ, Foster GA (1996) Acquisition of conditioned inhibition in rats is impaired by ablation of serotoninergic pathways. Eur J Neurosci 8: 415–423.
78. McNaughton N, Corr PJ (2004) A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. Neurosci Biobehav Rev 28: 285–305.
79. Wickens JR, Budd CS, Hyland BI, Arbuthnott GW (2007) Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. Ann N Y Acad Sci 1104: 192–212.
80. Hollis K (1984) The biological function of Pavlovian conditioning: the best defense is a good offense. J Exp Psychol Anim Behav Process 10: 413–425.
81. Heinz A, Mann K, Weinberger DR, Goldman D (2001) Serotonergic dysfunction, negative mood states, and response to alcohol. Alcohol Clin Exp Res 25: 487–495.
82. Horvitz JC (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96: 651–6.
83. Fadok JP, Dickerson TMK, Palmiter RD (2009) Dopamine is necessary for cue-dependent fear conditioning. J Neurosci 29: 11089–11097.
84. Beninger R, Mason S, Phillips A, Fibiger H (1980) The use of extinction to investigate the nature of neuroleptic-induced avoidance deficits. Psychopharmacology 69: 11–18.
85. Brischoux F, Chakraborty S, Brierley DI, Ungless MA (2009) Phasic excitation of dopamine neurons in ventral vta by noxious stimuli. Proc Natl Acad Sci U S A 106: 4894–4899.
86. Canli T, Lesch KP (2007) Long story short: the serotonin transporter in emotion regulation and social cognition. Nat Neurosci 10: 1103–1109.
87. Wyvell CL, Berridge KC (2000) Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. J Neurosci 20: 8122–8130.
88. Lex B, Hauber W (2010) The role of nucleus accumbens dopamine in outcome encoding in instrumental and pavlovian conditioning. Neurobiol Learn Mem 93: 283–290.
89. Heinz A, Goldman D, Jones DW, Palmour R, Hommer D, et al. (2000) Genotype influences in vivo dopamine transporter availability in human striatum. Neuropsychopharmacology 22: 133–139.
90. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104: 16311–16316.
91. Cools R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36: 98–113.
92. Frank MJ, Fossella JA (2011) Neurogenetics and pharmacology of learning, motivation, and cognition. Neuropsychopharmacology 36: 133–152.
93. Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, et al. (2010) Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. Psychol Med 40: 433–440.

94. Kahnt T, Park SQ, Cohen MX, Beck A, Heinz A, et al. (2009) Dorsal striatal-midbrain connectivity in humans predicts how reinforcements are used to guide decisions. J Cogn Neurosci 21: 1332–1345.
95. Bush R, Mosteller F (1955) Stochastic Models for Learning. New York, NY: John Wiley & Sons.
96. Narendra KS, Thathachar MAL (1974) Learning automata – A survey. IEEE Trans Syst Man Cybern 4: 323–334.
97. Tsypkin Y, Poznyak A (1972) Finite learning automata. Eng Cybern 10: 478–490.
98. Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13: 400–408.
99. Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8: 1704–1711.
100. Balleine BW (2005) Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav 86: 717–730.
101. Hall J, Parkinson JA, Connor TM, Dickinson A, Everitt BJ (2001) Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. Eur J Neurosci 13: 1984–1992.
102. Garcia J, Koelling R (1996) Relation of cue to consequence in avoidance learning. In: Houck LD, Drickamer LC, eds. Foundations of animal behavior: classic papers with commentaries. Chicago: University of Chicago Press. 374 p.
103. Beck A, Steer R, Brown G (1996) Manual for the Beck Depression Inventory-II. San AntonioTX: Psychological Corporation.
104. Beck AT, Epstein N, Brown G, Steer RA (1988) An inventory for measuring clinical anxiety: Psychometric properties. J Consult Clin Psychol 56: 893–897.
105. Spielberger C, Gorsuch R (1970) STAI manual for the State-trait anxiety inventory (" self-evaluation questionnaire"). Consulting Psychologists Press.
106. Montgomery S, Asberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134: 382.
107. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23: 56–62.
108. Shear MK, Bilt JV, Rucci P, Endicott J, Lydiard B, et al. (2001) Reliability and validity of a structured interview guide for the hamilton anxiety rating scale (sigh-a). Depress Anxiety 13: 166–178.
109. MacKay DJ (2003) Information theory, inference and learning algorithms. Cambridge, UK: Cambridge University Press.
110. Kass R, Raftery A (1995) Bayes factors. J Am Stat Assoc 90.