Don't Think, Just Feel the Music: Individuals with Strong Pavlovian-to-Instrumental Transfer Effects Rely Less on Model-based Reinforcement Learning

Miriam Sebold[1,2], Daniel J. Schad[1,3], Stephan Nebe[4], Maria Garbusow[1,2], Elisabeth Jünger[4], Nils B. Kroemer[4,5,6], Norbert Kathmann[2], Ulrich S. Zimmermann[4], Michael N. Smolka[4], Michael A. Rapp[3], Andreas Heinz[1], and Quentin J. M. Huys[7,8]

Abstract

■ Behavioral choice can be characterized along two axes. One axis distinguishes reflexive, model-free systems that slowly accumulate values through experience and a model-based system that uses knowledge to reason prospectively. The second axis distinguishes Pavlovian valuation of stimuli from instrumental valuation of actions or stimulus–action pairs. This results in four values and many possible interactions between them, with important consequences for accounts of individual variation. We here explored whether individual variation along one axis was related to individual variation along the other. Specifically, we asked whether individuals' balance between model-based and model-free learning was related to their tendency to show Pavlovian interferences with instrumental decisions. In two independent samples with a total of 243 participants, Pavlovian–instrumental transfer effects were negatively correlated with the strength of model-based reasoning in a two-step task. This suggests a potential common underlying substrate predisposing individuals to both have strong Pavlovian interference and be less model-based and provides a framework within which to interpret the observation of both effects in addiction. ■

INTRODUCTION

Pavlovian expectations of rewards or losses richly color and confound instrumental action choice. Background music is deployed in shops and restaurants to promote spending and specific choices, whereas stimuli associated with addictive substances are thought to perpetuate use and promote relapse. Individual variation in the nature of the underlying decision-making systems likely determines the strength of these effects.

Decision-making in humans and animals can be characterized along at least two axes, both of which are important for individual variation (Dayan & Berridge, 2014; Huys, Tobler, Hasler, & Flagel, 2014). The first axis concerns the distinction between model-free (MF) and model-based (MB) decision-making (Doll, Duncan, Simon, Shohamy, & Daw, 2015; Lee, Shimojo, & O'Doherty, 2014; Dezfouli & Balleine, 2013; Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Glascher, Daw, Dayan, & O'Doherty, 2010). The MF habit system learns through repeated experience, whereas the MB goal-directed system uses an internal model to prospectively reason about the value of actions. Computationally, MF decision-making relies on temporal difference learning: Values are learned through comparisons of estimated and actual received reward and updated with prediction errors. In MB reinforcement learning algorithms, the computation of values happens on the fly, integrating internal representations of state-action-reward probabilities and rewards (Sutton & Barto, 1998). Although MB decision-making is therefore computationally costly, MF decision-making is experientially demanding as changes have to be experienced multiple times for the iterative prediction error updates to change existing values. After an outcome devaluation (e.g., through satiation), the MB system can change preferences quickly, but the MF system cannot. Individual variation in the balance between MB and MF decisions, with a shift toward MF and away from MB learning, is associated with addictive and impulsive traits in animals (Huys et al., 2014; Everitt & Robbins, 2005), and a bias has been reported in conditions such as addiction and obsessive-compulsive disorder where behavioral preferences persist against explicit desires (Voon et al., 2014, 2015; Gillan et al., 2011, 2014; Sebold et al., 2014; Sjoerds et al., 2013).

The second axis concerns the distinction between instrumental and Pavlovian paradigms. In instrumental paradigms, actions a have values that depend on the presence of particular stimuli or situations s, leading to the valuation of stimulus–action pairs $V(s,a)$. In Pavlovian conditioning paradigms, stimuli s predict outcomes

[1]Charité-Universitätsmedizin Berlin, [2]Humboldt-Universität zu Berlin, [3]University of Potsdam, [4]Technische Universität Dresden, [5]Yale University School of Medicine, [6]The John B. Pierce Laboratory, New Haven, CT, [7]University of Zurich, [8]ETH Zürich

Journal of Cognitive Neuroscience 28:7, pp. 985–995
doi:10.1162/jocn_a_00945

independent of actions. These situations are described by action-independent stimulus values $V(s)$ (Dayan, Niv, Seymour, & Daw, 2006). Pavlovian values $V(s)$ influence actions in a variety of ways, including by eliciting approach/withdrawal to the stimulus s and by promoting or inhibiting the species-specific innate responses to s. They also have two distinct influences on instrumental processes in so-called Pavlovian–instrumental transfer (PIT) paradigms. Pavlovian stimuli influence the tendency to emit behavior generally (general PIT), with a stimulus predicting water for instance also enhancing responding for food, and they specifically increase choices of actions that lead to the outcome the Pavlovian stimulus is associated with (outcome-specific PIT). Individual variation in Pavlovian processes has again been related to addictive and compulsive traits (Garbusow et al., 2014; Flagel, Waselus, Clinton, Watson, & Akil, 2014; Flagel et al., 2011; Robinson & Berridge, 1993).

MB and MF systems have been shown to work in parallel in both instrumental and Pavlovian paradigms (Dayan & Berridge, 2014; Huys et al., 2014; Jones et al., 2012; Daw et al., 2011; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011; Daw, Niv, & Dayan, 2005; Killcross & Coutureau, 2003), leading to four values and many opportunities for complex interactions (Dayan & Berridge, 2014; Huys et al., 2014). For instance, outcome-specific PIT requires access to the specific nature of the outcome associated with the Pavlovian stimulus s. Computationally, this is by definition not contained in the MF value and, therefore, must depend on aspects of MB evaluation. On the other hand, devaluation of the outcome frequently fails to impact outcome-specific PIT (Eder & Dignath, 2015; Watson, Wiers, Hommel, & de Wit, 2014; Hogarth & Chase, 2011; Allman, DeLeon, Cataldo, Holland, & Johnson, 2010; Hogarth, Dickinson, & Duka, 2010; Corbit, Janak, & Balleine, 2007; Holland, 2004; Rescorla, 1994), suggesting computational mixtures, with MB processes for instance retrieving MF values that are resistant to devaluation. Indeed, possibilities for such complex interactions have been increasingly examined recently (Cushman & Morris, 2015; Huys et al., 2012, 2015; Guitart-Masip et al., 2012).

There are thus multiple paths toward the interaction between different valuation systems, and these are likely influenced by established individual variation both in terms of Pavlovian influences on choice and the balance away from MB decisions. We thus wanted to examine what the empirical, dominant pattern of covariation between MB/MF tradeoffs and Pavlovian influences on choice are in a healthy sample.

Specifically, we explored whether individual differences in PIT effects are associated with individual differences in the behavioral contribution of MB/MF learning in a separate instrumental choice task (Daw et al., 2011). We have previously observed increased PIT and reduced MB decisions in alcohol-dependent patients (Garbusow et al., 2014, 2015; Sebold et al., 2014) and hence expected PIT effects overall to be driven more by MF learning and to covary negatively with MB control. On the basis of these findings, we expected decreased MB but enhanced MF behavior in those participants with higher PIT effects. We aimed to test the described hypothesis in an exploration sample and replicate them in a secondary, demographically, and behaviorally distinct sample.

METHODS

Participants

At the time of analysis, a total of 267 participants were recruited as part of a longitudinal study on alcohol use disorder (LeAd study, www.Lead-studie.de, clinical trial numbers NCT01679145 and NCT01744834). The two-center study contains two separate projects. One project examines alcohol-dependent patients and age, sex, and education-matched healthy control participants. Because our hypotheses did not focus on alcohol dependence, we here examined healthy control participants only (n = 78). Data of 11 participants were excluded, two due to positive drug screenings, three due to technical issues, and another six due to poor task performance, leaving 67 participants (10 women, M_{age} = 43.07 years, SD_{age} = 11.02 years) for analyses. We first analyzed these participants and will therefore subsequently refer to them as the exploration sample. The second project examines 18-year-old male participants, representatively sampled from the local registry (n = 187). Data of two participants were removed due to technical issues, five due to positive drug screenings, two due to other exclusion criteria of the LeAd study (e.g., no alcohol intake in the past year), and two additional participants due to poor task performance, leaving 176 participants for analyses. Those participants were analyzed after the exploration sample, and we will thus henceforth refer to them as the replication sample. As the two samples differed profoundly in terms of demographics and behavior, this is a very stringent test. Both samples were examined for current and past psychiatric disorders using the Composite International Diagnostic Interview (Jacobi et al., 2013; Wittchen & Pfister, 1997). Exclusion criteria comprised a lifetime history of bipolar or psychotic disorder, current diagnosis of major depression, posttraumatic stress disorder, borderline personality disorder, obsessive-compulsive disorder, hypomania, generalized anxiety disorder, past and current substance dependencies other than nicotine, past and current neurological disorders, a history of severe head trauma, and current medication that affects the CNS.

Procedure

All participants first completed a PIT task and then the two-step task (Daw et al., 2011). Both tasks were programmed

using Matlab 2011 (version 7.12.0; The MathWorks, Natick, MA) with the Psychophysics Toolbox Version 3 (PTB-3; Brainard, 1997; Pelli, 1997). The two-step task and parts of the PIT task were performed inside an MRI scanner. The study was approved by local ethics committees. All participants gave written informed consent and were paid a fixed amount (€10/hr) plus an additional bonus contingent on their performance.

PIT Task

Participants underwent (1) instrumental training, (2) Pavlovian training, (3) PIT, and (4) a forced-choice task (see Garbusow et al., 2014). For description of each part, see Figure 1.

The task is notable in three features: in the use of approach; of both appetitive and aversive Pavlovian stimuli; and in that it contains instrumental stimuli for which either go or no-go yield more, but on average equal, reinforcement. The use of approach is motivated by the intuitive importance of maladaptive approach to drugs in addiction. By collapsing across equally valued go and no-go instrumental scenarios, it ensures that the PIT effect is not specific to active versus inactive responses. By including both gains and losses, it extracts Pavlovian conditioned stimuli (CS) effects that are related specifically to value independent of its sign.

Two-step Task

Each participant performed 201 trials of the two-step decision-making task described by Sebold et al. (2014; see Figure 2A). In each trial, participants had to perform an initial choice between two stimuli on a gray background. This choice then led to one of two second-stage options (either green or yellow) from which one stimulus

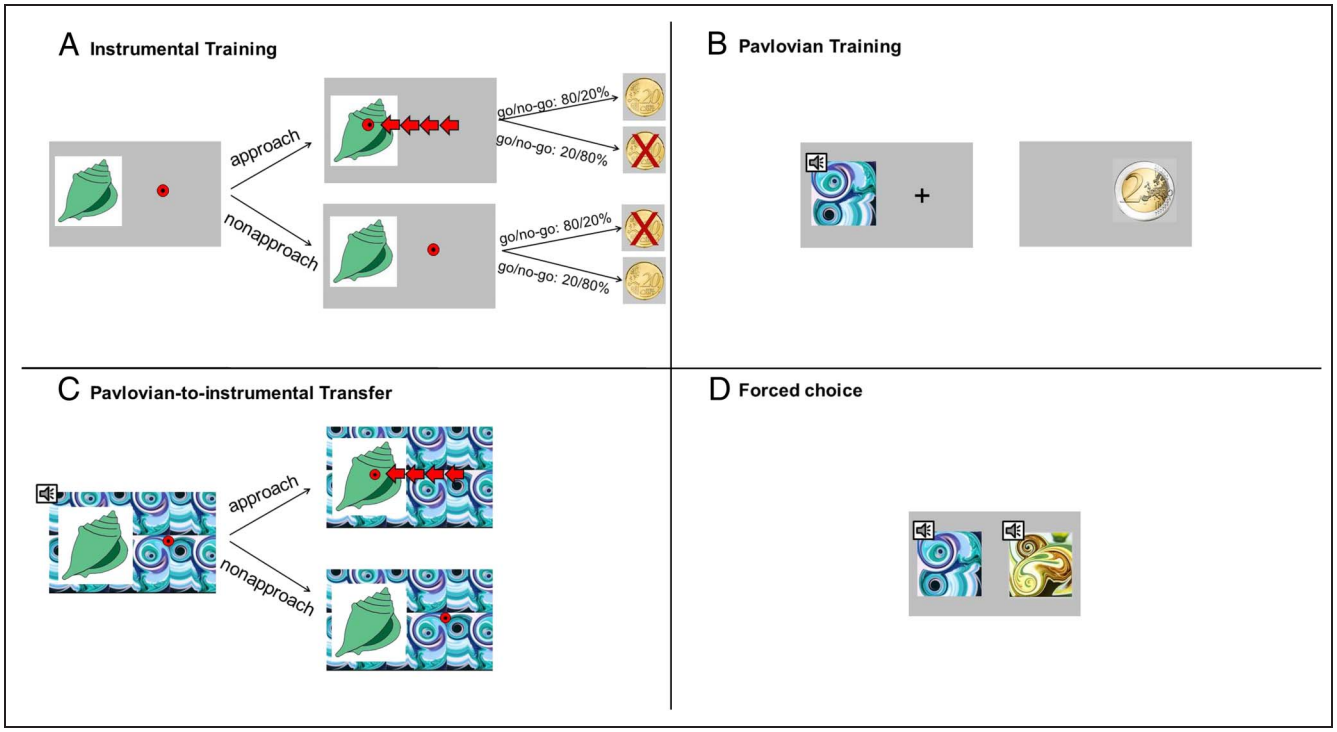

Figure 1. (A) Instrumental training: Participants were instructed to collect shells by repeated button presses after which they received probabilistic feedback. In "go trials", collection of a shell was monetarily rewarded in 80% and punished in 20% of trials, and vice versa if not collected. In "no-go trials", collection of a shell was monetarily punished in 80% and rewarded in 20% of the trials, and vice versa if not collected. A learning criterion for the instrumental training was enforced to ensure comparable task performance between participants (after a minimum of 60 trials, 80% correct choices over 16 consecutive trials). Participants performed the instrumental training until the learning criterion was met or for a maximum of 120 trials. (B) Pavlovian conditioning: At the beginning of each trial, participants saw a fractal-like stimulus accompanied by the sound of a tone (combined CS). After a delay of 3 sec, an unconditioned coin stimulus (US) was presented for another 3 sec. Participants were instructed to be attentive to the CS–US pairings. CS–US associations consisted of two CSs paired with images of +2/+1 EUR coins, one CS paired with 0 EUR, and two CSs paired with −1/−2 EUR, respectively. All participants completed 80 trials. (C) PIT: Each trial consisted of the presentation of one of the previously learned shells while both the auditory and visual CS from the Pavlovian conditioning were presented. Participants were instructed to perform the instrumental task again. Participants had 3 sec to respond. The intertrial interval was exponentially distributed ranging from 2 to 6 sec and a fixation cross displayed centrally. No feedback was presented, but participants were instructed that their choices would influence their final monetary outcome. There were 90 trials. (D) Forced choice task: Participants were presented with the two combined CS sequentially and asked to choose one. All possible CS pairings were presented three times in a randomized order. We used these data to verify acquisition of Pavlovian expectations and excluded participants for further data analyses (exploration sample $n = 6$, replication sample $n = 2$) if they did not perform better than chance in this part.

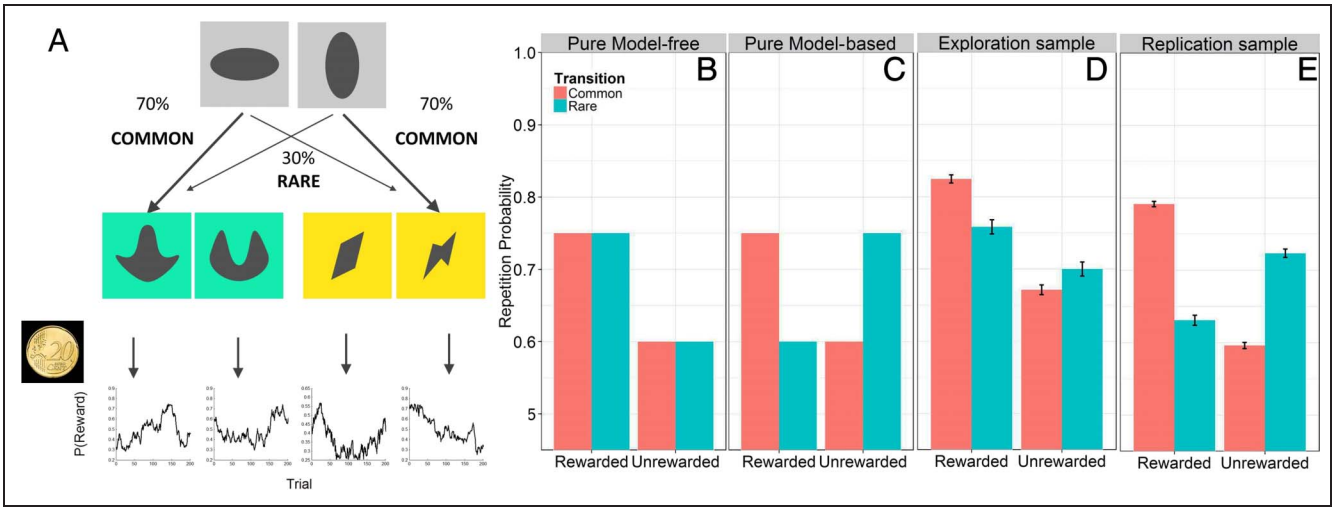

Figure 2. (A) The structure of the two-step task. In each trial, participants chose between two initial stimuli, leading them to a second stage (either green or yellow), at which they again had to make a choice. Each second-stage choice was probabilistically rewarded. These reward probabilities slowly changed over time. Each first-stage choice was frequently associated with a certain transition to the second stage (70% of all trials) but rarely associated with the opposing second stage (30% of trials). (B) MF decision-making does not consider transition frequencies. Stage 1 actions resulting in reward have a higher probability to be repeated than actions that did not end up being rewarded. Thus, MF decision-making predicts a main effect of reward. (C) Only MB decision-making takes transition probabilities into account. After a rewarded rare transition, the best chance of reaching that same rewarding second-stage stimulus again is to switch stimuli at the first stage and thereby use the frequent transition. Likewise, after a rare, unrewarded transition, the best chance of avoiding that same stimulus is to stay at this same first-stage stimulus, which commonly leads to the other, possibly rewarding second-stage stimuli. Both exploration (D) and replication (E) samples show a mixture of MB and MF choices.

had to be selected again. Crucially, the transition from first-stage choices to the specific second stage was probabilistic: Whereas one option on the first stage led frequently to the green second-stage option (70%) but rarely to the yellow second-stage option (30%), the other first-stage choice was associated with frequent yellow second-stage visits but rare green second-stage visits. At the second stage, participants were probabilistically rewarded with 20 cents or 0 cent (red cross superimposed on the 20-cent coin). To encourage participants to learn throughout the experiment, all four second-stage payoff probabilities changed slowly according to Gaussian random walks with reflecting boundaries at 0.25 and 0.75. We used the same random walk as in the original publication. In each stage, participants had 2 sec to perform their response. Variable intertrial intervals were drawn from an exponential distribution between 1 and 6 sec. Before starting the task, participants completed a training session with different random walks and a different stimulus set. Crucially, the training version was carefully translated from the version implemented by Daw et al. (2011). MB and MF decisions make distinct predictions on how reward and transition should influence first-stage behavior (Figure 2B and C).

Data Analysis

We first analyzed data from the exploration sample and subsequently validated our results with the replication sample. All regression analyses were conducted using generalized linear mixed-effects models implemented with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R programming language, version 3.1.2 (cran.us.r-project.org). For orthogonal contrasts in linear mixed-effects models, we used effect coding (−0.5/+0.5). Computational modeling was performed in Matlab 2012–2015 (versions 8.0–8.5).

PIT Task

All analyses focused on the PIT part (see Figure 1C), when participants had to perform a previously acquired response in the presence of Pavlovian stimuli.

The number of button presses in each trial was modeled as a Poisson distribution in a generalized linear mixed-effects model. In each trial, it was regressed on the nominal Pavlovian value of the CS in the background (−2, −1, 0, +1, +2). The model contained an additional nuisance variable to remove the influence of instrumental value (go/no-go) from the foreground stimuli. The within-subject factors (intercept, main effect of Pavlovian value, instrumental value, and their interaction) were treated as random effects across participants. Specific instrumental stimuli (shells) and Pavlovian stimuli (fractals-like) were taken as additional crossed random effects to control for item effects. We extracted individual regression coefficients for the CS stimuli (henceforth referred to as PIT slope) for further analyses. As the PIT slope histograms were bimodal, we clustered participants into two groups using a mixture of Gaussians fitted with expectation

maximization (mixtools package; Benaglia, Chauveau, Hunter, & Young, 2009). We also tested whether the PIT regression coefficients were significant in individual participants. However, these are for descriptive purposes only: As participants did not respond at all on some trails, button presses showed a zero inflation.

Two-step Task

We performed two sets of analyses. The first was a mixed-effects logistic (Otto, Skatova, Madlon-Kay, & Daw, 2015; Schad et al., 2014; Otto, Raio, Chiang, Phelps, & Daw, 2013) where first-stage choices (stay/switch) were regressed on the previous trial outcome and transition frequency (common or rare). Within-subject factors (intercept, main effect of reward, main effect of transition and their interaction) were taken as random effects across participants.

RTs. Knowledge of the transition frequency is only used when decisions are model-based, whereas in MF decisions common and rare trials are considered as equivalent. Thus, the difference between second-stage RTs after common versus rare transitions should reflect the level of involvement of MB control (Deserno, Huys, et al., 2015). We therefore repeated the above analyses, but using log-transformed second-stage RTs. Values two standard deviations below mean (0.5% of cases) were excluded from further analyses. This step did not influence the results. For visualization, MB RT effects were calculated from the individual difference between mean second-stage RTs after rare versus common transitions.

Computational model. We additionally fitted a reparameterization of the original Daw et al. (2011) reinforcement learning model to the data. It contains an MF parameter (β_{MF}) that weighs the contribution of an MF temporal difference learner and a parameter (β_{MB}) that weighs contributions by the MB learner, which uses the transition matrix as well as the reward contingencies. We imposed broad Gaussian priors (mean 0, variance 10) on all parameters, and results are based on maximum a posteriori parameter estimates. The model fitted better than chance in 75% (55/67) of the participants in the exploration sample and 72% (126/176) of the participants in the replication sample. Table 2 reports the estimated parameters of both samples. For inference, all parameters were transformed such that they were unbounded, and we retained these transformations to test correlations. None of the conclusions are affected by this transformation.

Relationship between PIT and Two-step Tasks

To test whether PIT effects were related to two-step performance, we added individual PIT slopes (as z-transformed variable) as a between-subject predictor in the binomial models of the two-step task and tested its interactions with the other fixed effects in the model.

For RT analyses, we performed linear mixed-effects regression with PIT slopes (z-transformed) and transition frequency as predictors for second-stage RTs.

In addition, we correlated individual MB (β_{MB}) and MF (β_{MF}) subject parameters from the computational model with PIT coefficients (Spearman correlation).

RESULTS

Exploration Sample: Choices

There was a significant group level PIT effect (fixed effect Pavlovian value, $p < .0001$; see Figure 3A) such that participants pressed more when there was a positive background CS and less when it was negative. Approximately half of the participants showed an individually significant effect (slope significantly positive in 63% 42/67 participants). The PIT slope was $b = 0.27$ on average (fixed-effect coefficient) and varied substantially across participants (random-effect $SD = 0.36$), suggesting large interindividual variation in the extent to which actions are controlled by Pavlovian stimuli, which is in line with previous research on PIT effects in humans (Garbusow et al., 2014; Prévost, Liljeholm, Tyszka, & O'Doherty, 2012).

In the two-step task, group level behavior reflected a mixture of MF and MB decision-making. There were both a significant main effect of reward ($p < .0001$) and a significant interaction between reward and transition ($p < .0001$; see Figure 2D).

To examine the relationship between PIT and the trade-off between MB and MF choices, we performed two tests. First, we entered individual PIT effects as additional regressors in the two-step logistic regression and tested (1) Reward × PIT slope and (2) Reward × Transition × PIT slopes interactions. Significant interactions would indicate that a relationship exists between the extent to which actions are influenced by Pavlovian values and MF versus MB learning, respectively. Individual PIT effects significantly interacted with MB decision-making (Reward × Transition × PIT slope: $p < .05$), but not with MF behavior (Reward × PIT slope, $p > .05$; see Table 1 and Figure 3B); as hypothesized, the association between PIT effects and MB learning was negative. Thus, participants who showed larger PIT effects were less model-based.

There was also a significant negative interaction between transition and PIT (transition × PIT, $p < .05$), indicating that participants with small PIT effects tended to stay more after common compared with rare trials. Although the transition itself does not play a role in either MB or MF system, the fact that those individuals who were less sensitive to it were more sensitive to Pavlovian CSs is in keeping with a shift away from MB learning.

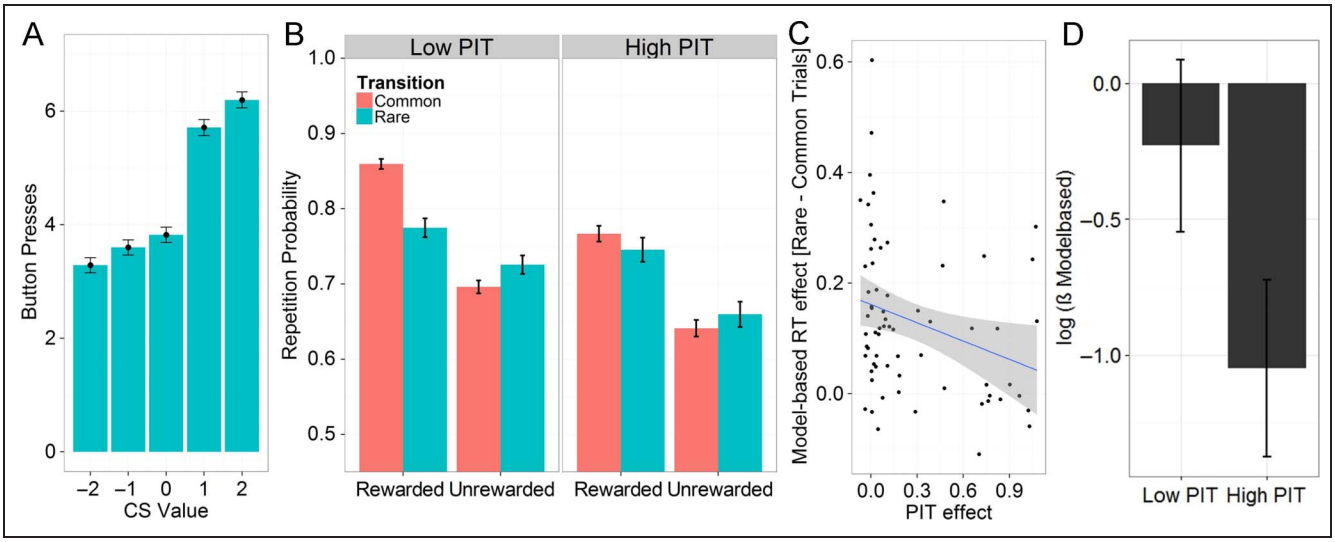

Figure 3. Results of the exploration sample. (A) Observed PIT effects. Button presses in the PIT task were strongly influenced by the value of the Pavlovian background (CS value). (B) Repetition probability as a function of reward and transition frequency in the exploration sample displayed separately for participants who show high and low PIT effects. Low PIT participants had a mean PIT effect of 0.03 (n = 41), whereas high PIT participants had an average PIT effect of 0.66 (n = 26). (C) Second-stage RT as a function of transition frequency covaried negatively with PIT effect: Participants who showed no PIT effect discriminated strongly between rare and common trials in their second-stage RTs, whereas participants who displayed large PIT effects did not show this discriminative second-stage RT behavior. (D) Estimates of the MB parameter β_{MB} displayed for participants who showed high and low PIT effects. Participants with high PIT values had lower β_{MB} parameter estimates.

Exploration Sample: Computational Modeling Results

Modeling analyses replicated these findings. There was a significant negative correlation between the weight given to MB choices, β_{MB}, and PIT coefficients ($r_{Spearman}$ = −.31, p < .01; see Figure 3D). There was no association between PIT and β_{MF} (p > .05).

Exploration Sample: RTs

Only the MB component has access to transition frequency. Hence, any difference in RTs between common and rare transitions should be related to the involvement of the MB system. RT differences between rare and common transitions correlated with β_{MB} ($r_{Spearman}$ = .49, p < .0001) but not with β_{MF} (p > .05) and with Transition × Reward effects ($r_{Spearman}$ = .59, p < .0001) but not with reward effects (p > .05), indicating that RT effects indeed reflect MB control. PIT effects again interacted negatively with transition (p < .01; Figure 4C). Participants with low PIT effects showed stronger transition effects on second-stage RTs and responded faster on common than rare trials.

Replication Sample: Choices

As in the exploration sample, there were significant PIT effects (fixed effect Pavlovian value, p < .0001; see

Table 1. Binomial Mixed-effects Results Testing the Influence of PIT Effects, Outcome of Previous Trials, and Transition of Previous Trial, upon Response Repetition for the Exploration and Replication Sample

	Exploration Sample		*Replication Sample*	
Coefficient	*Estimate (SE)*	*p*	*Estimate (SE)*	*p*
Intercept	1.36 (0.13)	<.0001*	0.96 (0.06)	<.0001*
Transition	0.24 (0.07)	.0006*	0.21 (0.04)	<.0001*
Reward	0.80 (0.09)	<.0001*	0.36 (0.04)	<.0001*
PIT slope	−0.17 (0.13)	.18	−0.03 (0.06)	.57
Transition × Reward	0.77 (0.17)	<.0001*	1.75 (0.14)	<.0001*
Transition × PIT slope	−0.13 (0.06)	.04*	−0.01 (0.04)	.75
Reward × PIT slope	−0.03 (0.09)	.73	0.06 (0.04)	.13
Reward × Transition × PIT slope	−0.41 (0.16)	.012*	−0.31 (0.14)	.03*

*p < .05.

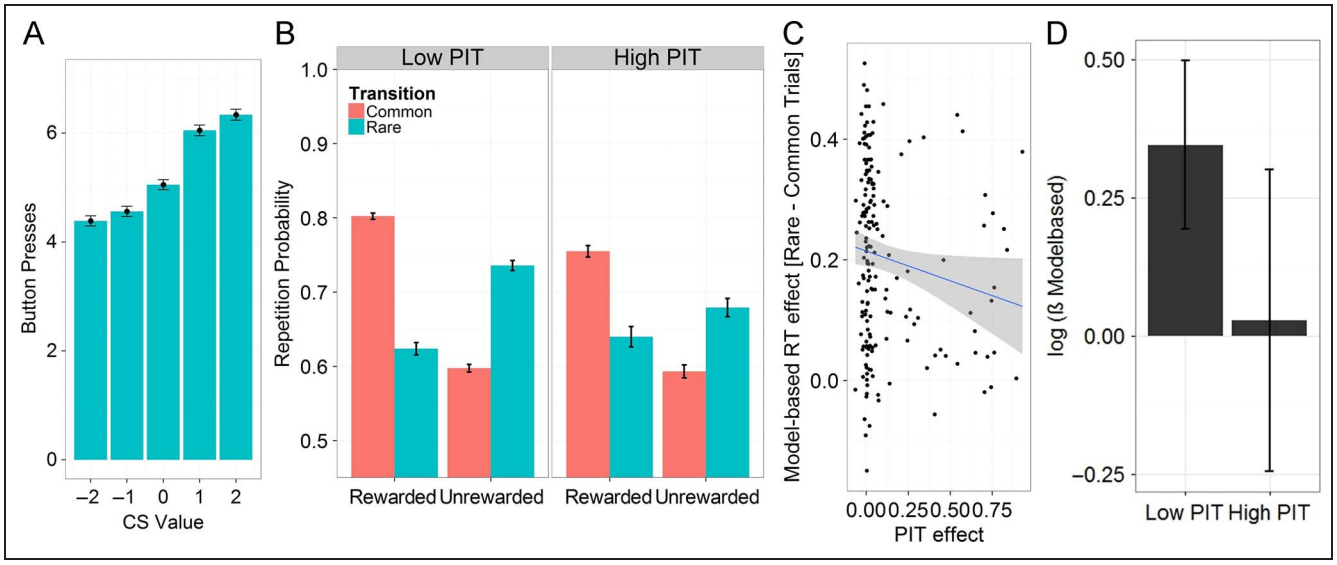

Figure 4. Results of the replication sample. (A) Observed PIT effects. Button presses in the PIT task are strongly influenced by the value of the Pavlovian background (CS value). (B) Repetition probability as a function of reward and transition frequency in the exploration sample separately displayed for participants who show high and low PIT effects according clustering of PIT effects as a mixture of Gaussians. Low PIT participants had a mean PIT effect of 0.008 (n = 130), whereas high PIT participants had an average PIT effect of 0.39 (n = 46). (C) Second-stage RT as a function of transition frequency negatively covaried with PIT effect: Participants who show no PIT effect discriminate strongly between rare and common trials in their second-stage RTs, whereas participants who display large PIT effects do not show this discriminative second-stage RT behavior. (D) Estimates of the MB parameter β_{MB} displayed for participants who show high and low PIT effects according to clustering of PIT effects as a mixture of Gaussians: Participants with high PIT values tended to have lower β_{MB} parameter estimates, even though this failed to reach statistical significance.

Figure 4A). However, the replication sample showed PIT effects less frequently (52/176 = 29% of participants), and the overall PIT slope (fixed-effect coefficient b = 0.12, random effect SD = 0.23) was numerically half the size of that in the exploration sample.

The two-step task again reflected a mixture of MF and MB decision-making with a significant main effect of reward ($p < .0001$) and a significant interaction between reward and transition ($p < .0001$). Results of interaction between PIT and all two-step parameters are outlined in Table 1. As in the exploration sample, individual PIT effects interacted with MB decision-making (Reward × Transition × PIT slope: $p < .05$), but not with MF behavior (Reward × PIT slope, $p > .05$; see Table 1 and Figure 4B). Again, the association between PIT effects and MB learning was negative, indicating that participants with large PIT effects used less MB behavior in the two-step task.

Of note, however, participants in the replication sample were younger (18 vs. 43.1 years on average) and, in keeping with previous results, were substantially more MB but less MF (Age × Reward × Transition, $p < .01$ and Age × Reward, $p < .0001$).

Replication Sample: Computational Modeling Results

There was no association between β_{MF} and PIT ($p > .05$), which mirrors the results from the regression analyses. However, the correlation between individual β_{MB} and PIT coefficients also failed to reach significance ($p > .05$). Upon visual inspection, participants with high PIT values tended to have lower β_{MB} values (Figure 4D). For exploratory purposes, we conducted an additional analysis among the high PIT effect group for whom the model fitted better than chance. Within this subgroup, PIT effects were negatively correlated with β_{MB} ($r_{Spearman} = -.37, p < .05$) but not with β_{MF} ($p > .05$).

Table 2. Estimates for All Parameters Shown as the Medians Plus Quartiles across Participants

	Exploration Sample							*Replication Sample*						
	β_{MB}	β_{MF}	ρ	β_2	α_1	α_2	λ	β_{MB}	β_{MF}	ρ	β_2	α_1	α_2	λ
25th percentile	0.09	1.29	0.32	1.7	0.34	0.33	0.36	0.64	0.76	0.17	1.7	0.26	0.39	0.23
Median	0.76	2.41	0.72	2.52	0.57	0.62	0.61	2.08	1.43	0.49	2.64	0.62	0.63	0.52
75th percentile	3.49	3.77	1.11	3.87	0.79	0.82	0.96	4.87	2.49	0.95	3.71	0.91	0.80	0.91

β_{MB} = MB component; β_{MF} = MF component; ρ = stickiness parameter indicating first-order perservation; β_2 = inverse temperature; α_1 = first-stage learning rate; α_2 = second-stage learning rate; λ = eligibility trace decay parameter.

Replication Sample: RTs

Analysis of the second-stage RTs also replicated the results of the exploration sample, with individual PIT effects showing a trend toward interacting negatively with transition ($p < .05$; Figure 4C and Table 2).

DISCUSSION

We examined the relationship between Pavlovian influences on behavior and the distinction between MB and MF choices. Across two independent and demographically diverse samples, we found that the extent to which Pavlovian values exerted control over behavior covaried negatively with MB decision-making in an independent task. In other words, participants whose decisions were strongly controlled by Pavlovian values also expressed decreased contributions of deliberative MB strategies. The same pattern was evident in RT analyses. Computational modeling analyses revealed equivalent direction of effects, as the MB parameter β_{MB} from a hybrid reinforcement learning model was negatively associated with PIT effects, although this association was only significant in one of the two samples.

The PIT paradigm we employed could theoretically allow for both outcome-specific and general PIT effects: The fact that the reward in the instrumental task and in the Pavlovian conditioning were both monetary suggests that outcome-specific PIT effects might be present. However, the parametric effect of CSs on behavior we observe clarifies that the value of the stimulus, not just its identity, is retrieved and influences choice. What we can say, then, is that the tendency to retrieve the value of a CS in PIT covaries negatively with MB reasoning in healthy populations. We therefore judge it strongly unlikely that the CS value retrieved would itself rely on MB processes and judge it more likely that it depends on MF ones. Such an interpretation is in accordance with recent work on individual variation in Pavlovian conditioning: Sign-trackers, who per definition express increased approach behavior toward conditioned cues, have stronger MF phasic dopaminergic signals (Flagel et al., 2011). Furthermore, they show less MB learning in that they are less sensitive to devaluation (Morrison, Bamkole, & Nicola, 2015) and Pavlovian extinction (Ahrens, Singer, Fitzpatrick, Morrow, & Robinson, 2016), and abolishing their MF learning through dopamine blockade does not uncover alternative MB reasoning (Flagel et al., 2011). Moreover, in humans, sign-trackers express increased PIT effects (Garofalo & di Pellegrino, 2015). As mentioned in the Introduction, in outcome-specific PIT the outcome must be explicitly accessed through a mental representation (a mental model) not available to the MF system and has hence been associated with the MB prospective system (Cartoni, Puglisi-Allegra, & Baldassarre, 2013; Dolan & Dayan, 2013; Clark, Hollon, & Phillips, 2012). Recent work has shown that the MB system can also access MF values (Cushman & Morris, 2015), which might explain the persistence of outcome-specific PIT after devaluation (Eder & Dignath, 2015; Watson et al., 2014; Corbit et al., 2007; Holland, 2004; Rescorla, 1994) and extinction (Rosas, Paredes-Olay, Garcia-Gutierrez, Espinosa, & Abad, 2010). However, such an interpretation of our data would have allowed even strongly MB participants to show strong PIT effects, which was not the case as it arose primarily in the absence of, or in conflict with, MB control.

In addition to a negative correlation with MB, we had also predicted a positive correlation between MF decision-making and (general) PIT effects, both because the two-step task measures a tradeoff between MF and MB (Doll, Bath, Daw, & Frank, 2016; Daw et al., 2011), but also because we had expected the strength of MF behavior in the two-step task to covary with the strength of Pavlovian MF conditioning and for that reason to promote general PIT. Against our expectations, we did not find a relationship between MF behavior and PIT, neither through regression analyses nor by analyzing the MF component from the computational model. This is likely because the task does not have much power to detect variation in the MF component, particularly separately from MB variation (cf. Doll et al., 2016). Most studies have found correlations with the MB but not with the MF component, including cognitive (Schad et al., 2014; Otto et al., 2013) and emotional (Otto et al., 2013) variables as well as pharmacological challenges (Worbe et al., 2015; Wunderlich, Smittenaar, & Dolan, 2012), brain stimulation (Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013; but see Smittenaar, Prichard, FitzGerald, Diedrichsen, & Dolan, 2014), and interindividual differences such as age (Eppinger, Walter, Heekeren, & Li, 2013) or psychiatric disorders (Sebold et al., 2014; Voon et al., 2014). Other tasks such as the probabilistic selection task may be more appropriate to specifically assess the MF system (Doll et al., 2016). Finally, it is worth noting that the reward effect in the one-step repetition probabilities is strongly influenced by the λ parameter in the model. This parameter directly determines how strongly a reward at the second step impacts on MF expectations at the first step. The MF weight β_{MF}, however, could also theoretically be large without such an effect, that is, for $\lambda = 0$ when a one-step repeat probability would show little reward effect. Hence, analyses of the reward-related repeat effects relate to aspects of the MF system more than to its overall behavioral dominance.

The study has some limitations. First, it is not entirely clear that other, more general mechanisms might have mediated the described association between both tasks. For instance, decreased MB performance and increased PIT effects might be caused by misunderstanding the instruction of either task. Specifically, we instructed all participants to rely on transition frequencies in the two-step task and to respond to the foreground stimuli in the PIT task (which interferes with PIT effects). Thus, those participants who showed decreased PIT effects and strong

MB control might have also been those who were more attentive to the instructions. A second limitation is that, at least in the replication sample, Pavlovian values tended to have comparably little influence on choice behavior and only a small number of participants showed PIT effects at all. Thus, the correlation between behaviors in both tasks is likely to be caused by a subset of participants only. Indeed when we correlated the MB parameter from the computational modeling with the PIT coefficients, the association became only significant when we limited our sample to participants with comparably high PIT effects. Moreover, we note that there were strong differences in the MF and MB component of the two-step task between the exploration and the replication sample. The samples differed very substantially by age, and there is strong evidence that age reduces MB behavior (Eppinger et al., 2013). As such, the pattern emerging across the two samples is strongly supportive of the findings in both individual samples that a reduction in MB tendencies covaries with increase PIT effects.

Third, across both samples, we found a significant main effect of transition. Thus, participants tended to stay more after common compared with rare trials, an effect that is neither obviously related to MF or MB accounts. Even though this effect has not been observed in the original study (Daw et al., 2011), several other studies have reported it. It is a small effect that becomes apparent in large sample sizes (Voon et al., 2014; Skatova, Chan, & Daw, 2013). Thus, null findings might be due to a lack of statistical power. However, we speculate that rare trials might be particularly salient and induce subsequent response behavioral shifts by reengaging MB controllers (Yasuda, Sato, Miyawaki, Kumano, & Kuboki, 2004).

There is accumulating evidence that in substance dependence and disorders of compulsivity PIT effects are increased (Garbusow et al., 2014, 2015; Hogarth, Field, & Rose, 2013; Glasner, Overmier, & Balleine, 2005) whereas MB control appears to be disrupted (Sebold et al., 2014; Voon et al., 2014). Moreover, MB neural signatures are reduced in high-impulsive individuals (Deserno, Wilbertz, et al., 2015), and impulsivity further seems to be associated with PIT effects (Garofalo & di Pellegrino, 2015). Our findings suggest a common underlying mechanism driving individual variation, possibly increasing the risk to develop substance dependence.

Acknowledgments

We thank the LeAD study teams in Dresden and Berlin for data acquisition. This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, FOR 1617; grants HE 2597/13-1, HE 2597/14-1, HE 2597/15-1, RA 1047/2-1, SM 80/7-1, ZI 1119/3-1, WI 709/10-1, SCHA 1971/1-2, HE 2597/13-2, HE 2597/14-2, HE 2597/15-2, RA 1047/2-2, SM 80/7-2, ZI 1119/3-2, WI 709/10-2).

Reprint requests should be sent to Miriam Sebold, Department of Psychiatry and Psychotherapy, Charite-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany, or via e-mail: miriam.sebold@charite.de.

REFERENCES

Ahrens, A. M., Singer, B. F., Fitzpatrick, C. J., Morrow, J. D., & Robinson, T. E. (2016). Rats that sign-track are resistant to Pavlovian but not instrumental extinction. *Behavioural Brain Research, 296,* 418–430.

Allman, M. J., DeLeon, I. G., Cataldo, M. F., Holland, P. C., & Johnson, A. W. (2010). Learning processes affecting human decision making: An assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. *Journal of Experimental Psychology Animal Behavior Processes, 36,* 402–408.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *_lme4: Linear mixed-effects models using Eigen and S4_*. R package version 1.1-7. Available at CRAN.R-project.org/package=lme4.

Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. S. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software, 32,* 1–29.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Cartoni, E., Puglisi-Allegra, S., & Baldassarre, G. (2013). The three principles of action: A Pavlovian–instrumental transfer hypothesis. *Frontiers in Behavioral Neuroscience, 7,* 153.

Clark, J. J., Hollon, N. G., & Phillips, P. E. (2012). Pavlovian valuation systems in learning and decision making. *Current Opinion in Neurobiology, 22,* 1054–1061.

Corbit, L. H., Janak, P. H., & Balleine, B. W. (2007). General and outcome-specific forms of Pavlovian–instrumental transfer: The effect of shifts in motivational state and inactivation of the ventral tegmental area. *European Journal of Neuroscience, 26,* 3141–3149.

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences, U.S.A., 112,* 13817–13822.

Daw, N. D., Gershman, S., Seymour, B., Dayan, P., & Dolan, R. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204–1215.

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

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective & Behavioral Neuroscience, 14,* 473–492.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Networks, 19,* 1153–1160.

Deserno, L., Huys, Q. J., Boehme, R., Buchert, R., Heinze, H. J., Grace, A. A., et al. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences, U.S.A., 112,* 1595–1600.

Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., et al. (2015). Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Translational Psychiatry, 5,* e659.

Dezfouli, A., & Balleine, B. W. (2013). Actions, action sequences and habits: Evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Computational Biology, 9,* e1003364.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80,* 312–325.

Doll, B. B., Bath, K. G., Daw, N. D., & Frank, M. J. (2016). Variability in dopamine genes dissociates model-based and model-free reinforcement learning. *Journal of Neuroscience, 36,* 1211–1222.
Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience, 18,* 767–772.
Eder, A. B., & Dignath, D. (2015). Cue-elicited food seeking is eliminated with aversive outcomes following outcome devaluation. *Quarterly Journal of Experimental Psychology, 69,* 574–588.
Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience, 7,* 253.
Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience, 8,* 1481–1489.
Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., et al. (2011). A selective role for dopamine in stimulus-reward learning. *Nature, 469,* 53–57.
Flagel, S. B., Waselus, M., Clinton, S. M., Watson, S. J., & Akil, H. (2014). Antecedents and consequences of drug abuse in rats selectively bred for high and low response to novelty. *Neuropharmacology*, *76(Pt B),* 425–436.
Garbusow, M., Schad, D. J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S. P., et al. (2015). Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addiction Biology*. doi:10.1111/adb.12243.
Garbusow, M., Schad, D. J., Sommer, C., Jünger, E., Sebold, M., Friedel, E., et al. (2014). Pavlovian-to-instrumental transfer in alcohol dependence: A pilot study. *Neuropsychobiology, 70,* 111–121.
Garofalo, S., & di Pellegrino, G. (2015). Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. *Frontiers in Behavioral Neuroscience, 9,* 163.
Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., et al. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry, 75,* 631–638.
Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry, 168,* 718–726.
Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595.
Glasner, S. V., Overmier, J. B., & Balleine, B. W. (2005). The role of Pavlovian cues in alcohol seeking in dependent and nondependent rats. *Journal of Studies on Alcohol, 66,* 53–61.
Guitart-Masip, M., Huys, Q. J., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *Neuroimage, 62,* 154–166.
Hogarth, L., & Chase, H. W. (2011). Parallel goal-directed and habitual control of human drug-seeking: Implications for dependence vulnerability. *Journal of Experimental Psychology Animal Behavior Processes, 37,* 261–276.
Hogarth, L., Dickinson, A., & Duka, T. (2010). The associative basis of cue-elicited drug taking in humans. *Psychopharmacology, 208,* 337–351.
Hogarth, L., Field, M., & Rose, A. K. (2013). Phasic transition from goal-directed to habitual control over drug-seeking produced by conflicting reinforcer expectancy. *Addiction Biology, 18,* 88–97.
Holland, P. C. (2004). Relations between Pavlovian–instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology Animal Behavior Processes, 30,* 104–117.
Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology, 8,* e1002410.
Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., et al. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences, U.S.A., 112,* 3098–3103.
Huys, Q. J. M., Tobler, P. N., Hasler, G., & Flagel, S. B. (2014). The role of learning-related dopamine signals in addiction vulnerability. *Progress in Brain Research, 211,* 31–77.
Jacobi, F., Mack, S., Gerschler, A., Scholl, L., Hofler, M., Siegert, J., et al. (2013). The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). *International Journal of Methods in Psychiatric Research, 22,* 83–99.
Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., et al. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science, 338,* 953–956.
Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex, 13,* 400–408.
Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron, 81,* 687–699.
McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *Journal of Neuroscience, 31,* 2700–2705.
Morrison, S. E., Bamkole, M. A., & Nicola, S. M. (2015). Sign tracking, but not goal tracking, is resistant to outcome devaluation. *Frontiers in Neuroscience, 9,* 468.
Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences, U.S.A., 110,* 20941–20946.
Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience, 27,* 319–333.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.
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. *Journal of Neuroscience, 32,* 8383–8390.
Rescorla, R. A. (1994). Transfer of instrumental control mediated by a devalued outcome. *Animal Learning & Behavior, 22,* 27–33.
Robinson, T., & Berridge, K. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews, 18,* 247–291.
Rosas, J. M., Paredes-Olay, M. C., Garcia-Gutierrez, A., Espinosa, J. J., & Abad, M. J. F. (2010). Outcome-specific transfer between predictive and instrumental learning is unaffected by extinction but reversed by counterconditioning in human participants. *Learning and Motivation, 41,* 150.
Schad, D. J., Jünger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A. H., et al. (2014). Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Frontiers in Psychology, 5,* 1450.

Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hagele, C., et al. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology, 70,* 122–131.
Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T., Penninx, B. W., et al. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry, 3,* e337.
Skatova, A., Chan, P. A., & Daw, N. D. (2013). Extraversion differentiates between model-based and model-free strategies in a reinforcement learning task. *Frontiers in Human Neuroscience, 7,* 525.
Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron, 80,* 914–919.
Smittenaar, P., Prichard, G., FitzGerald, T. H., Diedrichsen, J., & Dolan, R. J. (2014). Transcranial direct current stimulation of right dorsolateral prefrontal cortex does not affect model-based or model-free reinforcement learning in humans. *PLoS One, 9,* e86850.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
Voon, V., Baek, K., Enander, J., Worbe, Y., Morris, L. S., Harrison, N. A., et al. (2015). Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Translational Psychiatry, 5,* e670.
Voon, V., Derbyshire, K., Ruck, C., Irvine, M. A., Worbe, Y., Enander, J., et al. (2014). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry, 20,* 345–352.
Watson, P., Wiers, R. W., Hommel, B., & de Wit, S. (2014). Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite, 79,* 139–148.
Wittchen, H.-U., & Pfister, H. (1997). *DIA-X Interviews: Manual Für Screening-Verfahren Und Interview; Interviewheft Längsschnittuntersuchung (DIA-X-Lifetime); Ergänzungsheft (DIA-X-Lifetime); Interviewheft Querschnittuntersuchung (DIA-X-12 Monate); Ergänzungsheft (DIA-X-12 Monate); PC-Programm Zur Durchführung Des Interviews (Längs- Und Querschnittuntersuchung); Auswertungsprogramm*. Frankfurt am Main: Swets Test Service.
Worbe, Y., Palminteri, S., Savulich, G., Daw, N. D., Fernandez-Egea, E., Robbins, T. W., et al. (2015). Valence-dependent influence of serotonin depletion on model-based choice strategy. *Molecular Psychiatry*. doi:10.1038/mp.2015.46.
Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron, 75,* 418–424.
Yasuda, A., Sato, A., Miyawaki, K., Kumano, H., & Kuboki, T. (2004). Error-related negativity reflects detection of negative reward prediction error. *NeuroReport, 15,* 2561–2565.