Dissociating neural learning signals in human sign- and goal-trackers

Daniel J. Schad[1,2*], Michael A. Rapp[1], Maria Garbusow[2], Stephan Nebe[3,4], Miriam Sebold[2], Elisabeth Obst[3], Christian Sommer[3], Lorenz Deserno[5,6], Milena Rabovsky[7], Eva Friedel[2,8], Nina Romanczuk-Seiferth[2], Hans-Ulrich Wittchen[3], Ulrich S. Zimmermann[3,9], Henrik Walter[2], Philipp Sterzer[2], Michael N. Smolka[3,10], Florian Schlagenhauf[2,5], Andreas Heinz[2], Peter Dayan[11,12], Quentin J.M. Huys[13-16]

[1] Cognitive Sciences, University of Potsdam, Potsdam, Germany

[2] Dept. of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Berlin, Germany

[3] Dept. of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany

[4] Department of Economics, Zurich Center for Neuroeconomics, University of Zurich, Zurich, Switzerland

[5] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[6] Dept. of Psychotherapy and Psychosomatics, University of Leipzig, Leipzig, Germany

[7] Dept. of Psychology, Freie Universität Berlin, Berlin, Germany

[8] Berlin Institute of Health, Berlin, Germany

[9] Department of Addiction Medicine and Psychotherapy, kbo Isar-Amper-Klinikum, Munich-Haar, Germany

[10] Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

[11] Gatsby Computational Neuroscience Unit, University College London, London, UK

[12] Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[13] Center for Addictive Disorders, Hospital of Psychiatry, University of Zürich, Zürich, Switzerland

[14] Translational Neuromodeling Unit, University of Zürich and Swiss Federal Institute of Technology, Zürich, Switzerland

[15] Division of Psychiatry, University College London and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK

[16] Complex Depression Anxiety and Trauma Service, Camden and Islington NHS Foundation Trust, London, UK

* Corresponding author: Daniel J. Schad (danieljschad@gmail.com; ORCID linked to account on Manuscript Tracking System).

Abstract

Individuals differ in how they learn from experience. In Pavlovian conditioning paradigms, where cues predict reinforcer delivery at a different goal location, some animals—so-called sign-trackers—come to approach the cue, whereas others, called goal-trackers, approach the goal. In sign-trackers, model-free phasic dopaminergic reward prediction errors underlie learning, which renders stimuli 'wanted'. Goal-trackers do not rely on dopamine for learning and are thought to use model-based learning. We demonstrate this double dissociation in 128 male humans using eye-tracking, pupillometry and fMRI informed by computational models of sign- and goal-tracking. We show that sign-trackers exhibit a neural reward prediction error signal that is not detectable in goal-trackers. Model-free value only guides gaze and pupil dilation in sign-trackers. Goal-trackers instead exhibit a stronger model-based neural state prediction error signal. This model-based construct determines gaze and pupil dilation more in goal-trackers.

Main

Learning from reinforcements involves multiple processes with distinct computational, neural and behavioral signatures. Consider a simple classical Pavlovian conditioning paradigm, where a cue (conditioned stimulus, CS) becomes predictive of a reward (unconditioned stimulus, US) by being repeatedly presented before the US. In so-called model-free reinforcement learning, learning occurs via reward prediction errors (RPE [1]) which quantify the difference between the value of the US that actually arrives and a current prediction of that value made on the basis of the CS. Integration of the experienced RPEs allows the predictions made by the CS, called a 'cached value' to become accurate. An alternative way of learning involves building a model which has two components: a 'transition structure', which captures the probability that one stimulus is followed by another, and a 'reward structure', which captures the value associated with each particular stimulus. Learning the transition structure [2,3] can occur by integration of a different sort of so-called state prediction errors (SPE [4]) which quantify the difference between the stimulus that actually occurs and the probability of this event that was estimated on the basis of the previous stimulus. This type of learning does not conflate transitions and rewards, and is hence more flexible when one of these changes. However, it

is also computationally more costly to use models to make inferences because it requires the information from the transition and reward structures to be integrated on the fly [5]. Model-free learning has been suggested to underlie habits and model-based learning goal-directed decision-making [5–8].

Individual differences in the balance of these learning processes determine how and what we learn from our experiences. In turn, these influence how we interpret and react to new experiences, and as such may influence the development of mental illness after adverse events or substance use [9]. In rodent Pavlovian conditioning experiments with a discrete CS presented at a different location from the US, two broad categories of subjects can be differentiated: 'sign-tracking' animals, who approach the appetitive CS during conditioning and only subsequently go to the location of the US, and 'goal-tracking' animals, who come to approach the location of the US rather than the CS when the latter is presented. Furthermore, sign-trackers, but not goal-trackers will work to obtain the CS after learning [10].

The behavioral differences between sign- and goal-trackers have a number of revealing neural correlates. Sign-trackers learn from reward prediction errors (RPEs) coded in the activity of dopamine neurons [1] and evident in the phasic release of dopamine in the nucleus accumbens [11]. These RPE signals initially respond to the rewarding USs, but across learning, shift responding from the US towards the predicting CS. Indeed, sign-trackers depend on the dopaminergic signal to learn, as systemic dopamine blockade disables learning [10]. That sign-trackers will work to obtain the CS after learning suggests that the dopaminergic RPE underlies a form of learning which attributes incentive salience to the CS to make it wanted and thus to turn it into a motivationally relevant stimulus [2,10,12–15]. This is in line with model-free Pavlovian learning, where predictive value is cached and conflates stimulus identity and reward, rendering the CS 'rewarding' even though it itself lacks an affective consequence [16]. Such a value allows the CS to directly elicit Pavlovian approach or avoidance responses [17]. By contrast, for goal-trackers, phasic dopaminergic signals do not evolve

with learning as would be expected from a RPE learning signal, and learning is insensitive to dopamine blockade. This hints at model-based rather than model-free learning [2,18]. Put together, these results suggest a double dissociation, with sign-trackers being predisposed to dopaminergic model-free learning, and goal-trackers to non-dopaminergic model-based learning.

While human sign- and goal-tracking have been investigated using eye-tracking [19], their neural substrates have not yet been examined. Moreover, while detailed animal results demonstrate the neural systems underlying learning in sign-trackers, theoretical predictions about the computational and neural mechanisms [4] underlying learning in goal-trackers have not been fully tested to date [20].

We therefore administered a Pavlovian conditioning task during functional magnetic resonance imaging (fMRI) to 129 healthy human subjects. We hypothesized that the gaze direction during a specifically designed Pavlovian conditioning phase might parallel the behavioral responses seen in animals and allow us to separate humans into sign- and goal-trackers. We then examined the contribution of model-free and model-based learning to gaze and pupillary responses, to Pavlovian-Instrumental-Transfer (PIT) behaviors and to blood-oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI) signals. We did not explicitly manipulate state learning in our present task. Instead, model-based learning accounts predict trial-related changes in uncertainty and state prediction errors, which we investigate here. We found convergent evidence for a double dissociation, with sign-trackers relying on model-free, and goal-trackers on model-based learning.

Results

Subjects performed a Pavlovian conditioning task, in which visual-auditory CSs were deterministically paired with monetary reinforcements (Fig. 1a): in each of 80 trials, one of five CSs, consisting of fractal-like pictures and tones, was presented for three seconds on one side of the screen. This was followed by a blank screen with two fixation crosses. Then, one out of five possible USs, consisting of pictures of coins indicating a monetary win or loss (-2, -1, 0, +1, +2 Euros), was presented on the

other side of the screen. The conditioning task was the second part of a Pavlovian-instrumental transfer (PIT) task [21] consisting of four parts (Supplementary Fig. 1). Eye-tracking and fMRI recordings were obtained during Pavlovian conditioning.

To identify individual differences between sign- and goal-trackers we examined gaze responses to CS presentation [19]. Based on previous animal work [10,22], we studied a gaze index defined as the percentage fixation time on the CS minus on the US location, and regressed this gaze index on true CS value for each subject. Fig. 1b shows the distribution of regression coefficients. As sign-trackers approach appetitive CSs [10,15] and avoid aversive CSs [23], we defined sign-trackers as the upper tertile [15] of subjects with a positive influence of CS value on the gaze index (yellow in Fig. 1b, *Ns* = 43). Conversely, we defined goal-trackers as those subjects whose gaze approached appetitive US locations [10,15] and avoided aversive US locations (blue in Fig. 1b, *Ns* = 43). With respect to the timing of gaze responses, early responses to CS presentation often reflect orienting responses driven by visual salience [22,24] that are insensitive to CS value or learning. We therefore identified sign- and goal-trackers by analyzing the gaze index in the last second of CS presentation. Alternative analytical approaches to defining the groups result in similar patterns (see Supplementary Information).

To study signatures of sign-tracking, we examined how gaze was directed to the CS, the location of later US presentation, or the background, and how gaze was biased by CS value [10,19,22]. To this end, we performed repeated measures ANOVA with factors location (CS, US, background), CS value (-2 to +2 Euros), and time from CS onset (seconds 1, 2, and 3). For post-hoc tests, we used two-tailed t-tests of linear contrasts testing a linear effect or a linear interaction effect (contrast of linear fits; stronger increase in one condition than another) against zero. Moreover, we studied linear effects of trial number (trials 1-80; coefficients from linear regression analyses) using repeated measures ANOVA with the same factors. After initial orienting responses [22,24] insensitive to CS value (no evidence for the interaction CS value x location: $F(8, 2267) = 0.464$, $p = .882$, $\eta_p^2 = 0$, 90% CI [0 0.0004]), an influence of CS value on gaze emerged. During the third second of CS presentation, a

high CS value attracted gaze towards the CS (linear CS value: t_{2267} = 3.48, p < .001, b = 0.061, SE = 0.018, 95% CI = [0.027 0.096]; CS value x location x sec 1-3 of CS presentation: F(11, 1367) = 3.04, p < .001, η_p^2 = 0.004, 90% CI [0.001 0.007]), and away from the US location (t_{2267} = -1.78, p = .075, b = -0.031, SE = 0.018, 95% CI [-0.066 0.003]) and background (t_{2267} = -1.70, p = .089, b = -0.030, SE = 0.018, 95% CI [-0.065 0.005]; CS value x location: F(8,2267) = 4.458, p < .001, η_p^2 = 0.005, 90% CI [0.002 0.009]; c.f. Supplementary Figure 2; for exemplary trials see Supplementary Figure 3). This appeared to reflect learning as this CS value effect increased over trials for the CS location (t_{2675} = 2.47, p = .014, b = 0.008, SE = 0.003, 95% CI [0.002 0.015]), decreased for the US location (trend: t_{2675} = -1.77, p = .077, b = -0.006, SE = 0.003, 95% CI [-0.013 0.0006]), but there was no evidence for a change across trials for the background (t_{2675} = 0.70, p = .485, b = -0.002, SE = 0.003, 95% CI [-0.009 0.004]; CS value x location x trial: F(8,2675) = 2.94, p = .003, η_p^2 = 0.003, 90% CI [0.0008 0.006]). We summarized this effect in the gaze index [10,22] (Fig. 1c), which we analyzed using non-parametric bootstrapping (1,000,000 case resamples and bias-corrected adjusted confidence intervals) with two-tailed statistical testing. The gaze index became increasingly biased towards the higher value CS (linear fit: $p_{bootstrap}$ < .05, b = 0.009, SE = 0.005, 95% CI = [0.001 0.022]), with the impact of value increasing over trials (interaction of linear fits: CS value x trial number, $p_{bootstrap}$ < .05, b = 0.0015, SE = 0.0008, 95% CI = [0.00001 0.0032]).

Hence, there appeared to be an eye-tracking signal in humans analogous to the behavioral sign-tracking response in animals [19,22]. We examined individual variation in this measure between sign- and goal-trackers. For this, we tested linear contrasts within each group (linear fits and interactions between linear fits indicating a stronger increase in one condition than another). The group of sign-trackers fixated win-predictive CSs more than loss-predictive CSs (linear fit of CS value; $p_{bootstrap}$ < .001, b = 0.059, SE = 0.011, 99.9% CI = [0.039 0.129]; Fig. 1d-f), and fixated aversive CSs progressively [2,3,17] less over time, instead increasingly fixating the US-location when anticipating aversive USs (linear fit of trial number for aversive CSs; $p_{bootstrap}$ < .001, b = −0.013, SE = 0.005, 99.9% CI = [−0.037 −0.001]; linear CS value x linear trial number: $p_{bootstrap}$ < .05, b = 0.003, SE = 0.002, 95% CI = [0.0003

Inf]). Conversely, the group of goal-trackers fixated the US location more for appetitive than aversive USs and vice versa for the CS (linear fit of CS value; $p_{bootstrap} < .001$, $b = -0.036$, $SE = 0.004$, 99.9% CI = [−0.053 −0.027]), and did so progressively over time.

So far, the definition of goal-tracking is simply converse of the definition of sign-tracking and hence not an independent measure. We therefore looked for more specific signatures of model-based learning that should uniquely characterize goal-trackers. Gaze is known to reflect uncertainty about the consequences of a stimulus independently of value [25,26]. We reasoned that this should reflect the learning of the model through SPEs, which are larger when there is more uncertainty about the predicted stimulus identity independently of its associated reward. Hence, this would predict that the attraction of gaze to the CS should simply reduce over the course of the experiment, and this should be more prominent amongst goal- than sign-trackers. The last second of CS presentation indeed revealed a strong effect of trial in addition to the above value effects. Gaze was strongly focused on the initially uncertain CS location early on, but continuously drifted away from the CS (linear trial effect: $t_{410} = -8.62$, $p < .001$, $b = -0.008$, $SE = 0.001$, 95% CI [-0.010 -0.007]) towards the US-location ($t_{410} = 3.21$, $p = .001$, $b = 0.003$, $SE = 0.001$, 95% CI [0.0012 0.0050]) and the background ($t_{410} = 5.41$, $p < .001$, $b = 0.005$, $SE = 0.001$, 95% CI [0.003 0.007]; trial x location: $F(2,410) = 37.93$, $p < .001$, $\eta_p^2 = 0.013$, 90% CI [0.008 0.018]). As a result, the gaze index was biased away from the CS towards the US-location (i.e., it decreased) with increasing trial number ($p_{bootstrap} < .001$, $b = -0.011$, $SE = 0.002$, 99.9% CI = [−0.017 −0.003]; Fig 1c).

To directly test whether this reflected the reduction in uncertainty over the course of training, we implemented computational models assuming gaze is controlled either by trial-by-trial uncertainty from a model-based learning system or by Pavlovian responses to model-free trial-by-trial CS value (c.f. Methods). We computed BIC values for both models for each subject, and performed a repeated measures ANOVA with factors model (model-free value versus model-based uncertainty) and group (sign- versus goal-trackers). We performed post-hoc tests using two-tailed t-tests of the

difference in BIC-values between models with each group of sign- versus goal-trackers separately. We found that in goal-trackers, the gaze index was best explained by the uncertainty-based model (t_{84} = −3.28, p = .002, ΔBIC = −3.73, SE = 1.14, 95% CI = [-6.00 -1.47]; see Fig. 1g), suggesting state uncertainty drives gaze in goal-trackers. The evidence in sign-trackers' gaze, to the contrary, was significantly shifted towards the value-based model (model x group: $F(1, 84) = 6.87$, p = .010, η_p^2 = 0.038, 90% CI = [0.005 0.097]), but provided no statistical evidence supporting one model over the other (t_{84} = 0.43, p = .672, ΔBIC = 0.48, SE = 1.14, 95% CI = [-1.78 2.75]). Additional modeling, allowing for dual control where value-based ($\omega = 0$) and uncertainty-based ($\omega = 1$) learning systems within each subject are combined via a weighting parameter (ω), suggested that goal-trackers relied strongly on uncertainty- or model-based control (ω_{Mean} = 0.84, ω_{SD} = 0.18), whereas sign-trackers seem to use a mixture of value- and uncertainty-based systems (ω_{Mean} = 0.48, ω_{SD} = 0.24; we tested the group-difference in the ω parameter using two-tailed non-parametric bootstrapping: $p_{bootstrap}$ < .001, b = 0.36, SE = 0.05, 99.9% CI = [0.20 0.50]; Fig. 1h; c.f. Methods). Hence, examining changes in how individuals freely chose to gaze at a CS or a US allowed us to distinguish two groups of subjects which appear to rely on different computational mechanisms for learning.

The pupil is known to dilate in response to uncertainty, proposedly reflecting noradrenergic arousal signals in nucleus coeruleus and associated sites [27]. Moreover, the pupil dilates during learned anticipation of rewards relative to losses or neutral outcomes, putatively reflecting Pavlovian motivation or arousal signaled in noradrenaline and elicited by an anticipatory dopamine response [28]. As such, we expected to see a similar distinction between goal- and sign-trackers as we saw in gaze control. We focused on the last second before US onset to avoid luminance effects due to the stimuli, to avoid temporal transients, and because incentive salience is thought to peak just prior to US onset [29]. We first asked whether pupil size was driven by uncertainty versus CS value and again found a double dissociation. To this end, we performed repeated measures ANOVA with factors trial number (3-8 versus 9-16), CS value (-2 to +2 Euros), time since CS onset (seconds 1 to 6), and group

(sign- versus goal-trackers). Post-hoc tests for second six after CS onset were performed using t-tests, testing effects in each group against zero. In these contrasts, a simple between-group t-test between sign- and goal-trackers provides evidence for an interaction (stronger increase in one group than another) because it is a contrast of linear fits. In goal-trackers, average pupil size decreased from the beginning to the end of conditioning, consistent with the decrease in uncertainty occasioned by learning (effect of trials [3-8 vs. 9-16]: $t_{140} = -2.29$, $p = .023$, $b = -0.055$, $SE = 0.024$, 95% CI = [-0.102 -0.008]). No effect of trial number was observed in sign-trackers ($t_{140} = 0.83$, $p = .405$, $b = 0.020$, $SE = 0.025$, 95% CI = [-0.028 0.069]), with a significant group difference ($F(1, 140) = 4.84$, $p = .030$, $\eta_p^2 = 0.001$, 90% CI = [0 0.003]; Fig. 2a+c). A different signature was visible in sign-trackers' pupil size. Here, the pupil was dilated by the expectation of wins compared to neutral outcomes or losses in the second half of the experiment (trials 9-16), reflecting a value-based pupil response (linear CS value effect: $t_{1314} = 2.89$, $p = .004$, $b = 0.521$, $SE = 0.180$, 95% CI = [0.167 0.874]). This linear CS value effect developed from the beginning to the end of conditioning (significant increase; $t_{638} = 2.93$, $p = .004$, $b = 0.339$, $SE = 0.116$, 95% CI = [0.112 0.566]) reflecting learning of CS value. In goal-trackers, this was not observed: there was no evidence that CS value influenced pupil size (trials 9-16: $t_{1314} = -1.59$, $p = .112$, $b = -0.280$, $SE = 0.176$, 95% CI = [-0.625 0.066]; group difference: $t_{638} = 3.52$, $p < .001$, $b = 0.285$, $SE = 0.081$, 95% CI = [0.126 0.444]). Pupil dilation in goal-trackers hence appeared to reflect uncertainty [27], while it was driven by CS value in sign-trackers [28]. We next asked whether these could again be mapped onto model-based and model-free learning by studying BIC values for each model, computed across all individual subjects for each group of sign- and goal-trackers. In goal-trackers, pupil dilation was best accounted for by model-based uncertainty (model-based uncertainty: $BIC = 8023.6$; model-free CS value: $BIC = 8032.4$; $\Delta BIC = 8.81$; Fig. 2d) and explained the reduction in pupil size across trials seen in goal-trackers only (Fig. 2e). In sign-trackers, to the contrary, pupil dilation was best accounted by model-free CS value (model-based uncertainty: $BIC = 7460.1$; model-free CS value: $BIC = 7457.1$; $\Delta BIC = -2.93$; Fig. 2d) and this model was able to capture the continuous increase of the CS value effect across trials seen in sign-trackers only (Fig. 2f).

Hence there was again a double dissociation: pupil size reflected model-based uncertainty about upcoming states in goal-trackers, while it reflected model-free value in sign-trackers.

We next attempted to validate the distinction between sign- and goal-trackers in measures independent from eye-tracking. At a behavioral level, we examined two independent predictions. First, CSs are thought to acquire incentive salience [12–14] in sign- but not in goal-trackers [10,15], and to elicit Pavlovian-instrumental transfer (PIT) only in sign-trackers [19]. In PIT, appetitive CSs enhance and aversive CSs reduce instrumental approach [21]. The behavioral paradigm employed here contained a PIT phase, in which Pavlovian CSs were presented in the background of the instrumental task; no outcomes were presented but subjects were instructed that outcomes would count towards their reimbursement (Supplementary Fig. 1). We computed the PIT effect for each individual subject as the linear fit of Pavlovian CS value on the number of button presses, and used non-parametric bootstrapping to test the directed hypothesis (one-tailed) that the PIT effect is stronger and more frequently individually significant (tested using individual t-tests) in sign- than in goal-trackers. We found that the PIT effect was stronger ($p_{bootstrap} < .05$, $b = 0.49$, $SE = 0.26$, 95% CI = [0.09 Inf]; Fig. 3a,b) and more frequently significant at an individual level ($p_{bootstrap} < .05$, $b = 15.8$, $SE = 8.3$, 95% CI = [1.6 Inf]; Fig. 3a, inset) in sign-trackers than in goal-trackers, suggesting that the CS acquired incentive salience and elicited Pavlovian approach and avoidance behavior more in sign-trackers. Second, while sign- and goal-trackers learn differently, they should learn the Pavlovian values equally well. Our paradigm also contained a phase in which subjects were forced to choose the better amongst pairs of CSs, and, as expected, sign- and goal-tracker performance was excellent and not statistically different (goal-trackers: 97.8% correct, $SD = 9.2$; sign-trackers: 95.2% correct, $SD = 14.0$; group difference: $p_{bootstrap} > .1$, $b = 2.6$, $SE = 2.6$, 95% CI = [-1.8 8.6]; see also Supplementary Information, Supplementary Figure 4).

We finally turned to neuroimaging to more directly examine the nature of the learning signals in the two groups. Animal sign- but not goal-trackers have been shown to exhibit a temporal difference

reward prediction error (RPE) response in NAc dopamine concentrations during Pavlovian conditioning [10]. Such RPE signals in human ventral striatum can be measured with fMRI [30]. We computed trial-by-trial temporal difference RPEs for CSs and USs using a simple reinforcement learning model (Supplementary Information). The temporal difference RPE regressor was used in a linear model with factor group (sign- versus goal-trackers) and covariate testing site to explain the NAc BOLD response. An ANOVA with the factor group was used to test the difference in the RPE signal between sign- and goal-trackers, and one-sample t-tests were used to test whether the RPE signal was larger than zero within each group separately. Family-wise error (FWE) correction was used to control the peak-voxel effect for multiple tests associated with multiple voxels within the a priori volume of interest [VOI] in the NAc. The RPE explained a significant amount of variance in the NAc BOLD response in sign-trackers (small volume corrected [SVC] in NAc VOI: t_{75} = 3.05, p_{FWE} = .025, [12 6 -14]) but there was no evidence for such an effect in goal-trackers (t_{75} = 1.58, SVC p_{FWE} = .398), with a significant group difference ($F(1,75)$ = 10.88, SVC p_{FWE} = .026, [12 6 -14], η_p^2 = 0.122, 90% CI = [0.031 0.242]; Fig. 4a-d).

The RPE signal is evident in conditioning involving wins, but can be less clear for losses, which may even involve inverted RPE or salience signals [31]. We therefore repeated analyses testing the RPE for wins (0€, +1€, +2€) and losses (0€, -1€, -2€) separately. For this analysis, we extracted the average appetitive or aversive RPE BOLD signal within the a prior NAc VOI for each subject, and performed one-sample t-tests of the hypotheses that the appetitive RPE signal in sign-trackers is larger than zero, and larger than in goal-trackers. Results for the appetitive RPE involving wins were in line with the overall findings, namely a NAc BOLD RPE response in sign-trackers (t_{38} = 2.15, p = .019, b = 0.087, SE = 0.040, 95% CI = [0.019 Inf]; Fig. 5a+e), but no NAc RPE response in goal-trackers (Fig. 5a+b, t_{38} = -0.04, p = .516, b = -0.002, SE = 0.042, 95% CI = [-0.072 Inf]; group difference: t_{76} = 1.53, p = .065, b = 0.089, SE = 0.058, 95% CI = [-0.008 Inf]; Supplementary Fig. 5). The aversive RPE involving losses, however, did not elicit BOLD responses (p > .1; c.f. Supplementary Information).

RPE(-like) signals are found also in other brain regions such as ventral tegmental area/substantia nigra (VTA/SN), dorsal striatum (caudate and putamen), vmPFC, and amygdala, which can be dissociated from the NAc RPE signal in specifically designed tasks. Whether these signals are also selectively expressed in sign-trackers is currently unknown (for some evidence see [32]). We tested whether the difference in RPE signals between sign- and goal-trackers is also present in these other brain regions. We performed repeated measures ANOVAs with factors VOI (VTA/SN, Caudate, Putamen, vmPFC, amygdala) and group (sign- versus goal-trackers). For post-hoc tests, we used one-tailed t-tests to test the hypothesis that the RPE signal in each group was larger than zero. For our a priori analysis involving wins and losses, across these VOIs we found significant RPE responses in sign-trackers (t_{76} = 1.89, p = .031, b = 0.035, SE = 0.019, 95% CI = [0.004 Inf]), and these were stronger than in goal-trackers throughout (F(1,76) = 4.18, p = .044, η_p^2 = .01, 90% CI = [0.00 0.03]; RPE signal in goal-trackers: t_{76} = -1.00, p = .322, b = -0.019, SE = 0.019, 95% CI = [-0.049 Inf]). There was no evidence that the group-difference differed between VOIs (F(3,241) = 1.07, p = .363, η_p^2 = 0.003, 90% CI = [0 0.023]), indicating no reliable difference across regions. The same pattern was also true for the analysis involving wins only, where again sign-trackers showed a RPE signal (t_{76} = 2.38, p = .010, b = 0.080, SE = 0.034, 95% CI = [0.024 Inf]), which was stronger than in goal-trackers (F(1,76) = 5.40, p = .023, η_p^2 = 0.014, 90% CI = [0.001 0.039]), where goal-trackers showed no evidence for a RPE signal (t_{76} = -0.90, p = .369, b = -0.030, SE = 0.034, 95% CI = [-0.086 Inf]). We explored appetitive RPE signals in individual VOIs (Fig. 5a,c-f; Supplementary Figure 6). We found that appetitive RPE signals in sign-trackers were stronger than in goal-trackers in several VOIs (VTA: p = .038; vmPFC: p = .032; Putamen: p = .121; Caudate: p = .058; Amygdala: p = .004), and that only the effect in the Amygdala (p = .020), but not in the other VOIs (p > .1), survived correction for the multiple exploratory tests. However, these differences should be interpreted with caution given there was no evidence that VOIs differed. One problem with fMRI analyses of Pavlovian learning is that the correct learning rate is unknown and cannot be estimated easily from behavior [33]. However, the differences between sign- and goal-trackers were consistent across a range of different values

for the learning rate: in our a priori analysis of wins and losses, averaged across all VOIs the group-difference was significant for learning rates 0.1 (p = .034), 0.2 (p = .039), 0.3 (p = .046), 0.4 (p = .048), 0.5 (p = .0495), and 0.6 (p = .04998). A similar pattern was present when analysing wins only, but not when analysing losses only (see Supplementary Information; including Supplementary Fig. 7).

While theoretical accounts and the results so far suggest that goal-trackers may rely on model-based learning [2,3], comparatively little data exists [20] on the learning processes and neurobiological mechanisms in goal-trackers. Our computational account of model-based learning incorporates incremental updates to state expectations through SPEs [4]. Human fMRI results have previously shown SPEs to be represented in the intra-parietal sulcus and in lateral PFC [4]. Hence, if goal-trackers learn through model-based mechanisms, we expect more prominent SPEs in these areas in them than in sign-trackers. We tested whether SPE signals were different from zero within each group using two-tailed t-tests. Moreover, we performed repeated measures ANOVA with factors VOI (intraparietal sulcus and lPFC) and group (sign- and goal-trackers). A post-hoc two-sample t-test was used to test whether the SPE BOLD signal differed between groups within the intraparietal sulcus. We found SPE signals in both intraparietal sulcus and lPFC (Fig. 6a), which were significant in both goal-trackers (t_{76} = 6.44, p < .001, b = 1.165, SE = 0.181, 95% CI = [0.804 1.53]) and sign-trackers (t_{76} = 4.94, p < .001, b = 0.894, SE = 0.181, 95% CI = [0.534 1.25]; also see Supplementary Figure 8), consistent with the behavioral signatures showing a model-based component in both groups. However, in the intraparietal sulcus this SPE signal was stronger in goal- than in sign-trackers (t_{67} = 2.12, p = .038, b = 0.564, SE = 0.266, 95% CI = [0.034 1.095]; interaction group x VOI: F(1, 76) = 5.34, p = .024, η_p^2 = 0.033, 90% CI = [0.002 0.092], Fig. 6a,b, Supplementary Fig. 9). As previous work [4] has also shown this area to relate to behavior, this difference may underlie goal-trackers' stronger reliance on model-based Pavlovian learning in the current task.

Finally, if it is true that the imaging and eye measures relate to the same learning processes, then the learning parameters estimated from the two modalities should not be too dissimilar. Keeping

the difficulties in estimating learning rates in mind, we nevertheless found that the uncertainty-based signal in goal-trackers yielded a state learning rate of η_{pupil} = 0.19, which was well in line with the learning rate of η_{fMRI} = 0.15 used for fMRI analyses. The value signal in sign-trackers yielded a learning rate for value of α_{pupil} = 0.06, which was comparable with the learning rate showing the strongest NAc RPE signal in model-based fMRI analyses (α_{fMRI} = 0.05; see Supplementary Fig. 7).

Discussion

In summary, we found a double dissociation between model-free and model-based Pavlovian learning systems in human sign- versus goal-trackers. As a key neurobiological finding, the model-free RPE teaching signal in the (ventral) striatum was present in human sign-trackers, but was not detectable in goal-trackers, suggesting that, as in animals [10], only sign-trackers rely on model-free RPE signals [1,11] for learning. Model-free learning assigns value to the CS, which turns it into a motivationally relevant stimulus that elicits approach and avoidance responses in its own right. We here found that the value of the CS elicited approach and avoidance responses during conditioning, as measured in influences of CS value on gaze and pupil size during conditioning, and on Pavlovian-instrumental transfer (PIT). Sign-trackers thus seem to rely on a model-free learning system that uses RPE signals to attribute incentive salience to the CS. In goal-trackers, to the contrary, gaze and pupil size related to model-based uncertainty. This was accompanied by a stronger model-based SPE signal in intraparietal sulcus. Goal-trackers also showed less PIT effects [17,34] suggesting that the CSs did not acquire the same motivational properties. Goal-trackers thus seem to rely on model-based reinforcement learning to predict the identity of upcoming US states from the CS. Of note, the neural distinction does not only show a dissociation of the (ventral) striatal versus (intra-) parietal brain regions for learning in sign- and goal-trackers, but also of the computational mechanisms driving learning in each. Furthermore, for learning success as assessed in a forced choice task, we did not find a difference between sign- and goal-trackers. Group differences, as in animals [10], may hence not reflect differences in learning ability, but rather indicate different mechanisms or systems

underlying learning. Taken together, eye-tracking, pupillometry, behavioral PIT responses and fMRI consistently reveal a double dissociation between model-free RPE learning mediating incentive salience attribution in sign-trackers versus model-based SPE learning guiding uncertainty-based selection in goal-trackers.

Strikingly, the RPE signal was not only present in the ventral striatum, but extended throughout a broad affective area including the dorsal striatum (Putamen, Caudate), VTA, Amygdala, and vmPFC in the signtrackers, while we found no evidence for RPE signals in these areas in the goal-trackers. Early theories (e.g. [35]) had conceived of the RPE signal as a dopaminergic teaching signal with wide applicability to be broadcast throughout the brain. The fact that the RPE signal only behaves as a teaching signal in sign-trackers, and only seems to be broadcast widely in them, is consistent with such an account. Given the extensive literature on the relationship between RPEs and phasic dopaminergic signals [1,10,36], it is highly likely that the RPE signal in sign-trackers is dopaminergic.

In model-free learning, the predictive value computed by iteratively adding up RPE signals is assigned directly to the CS. This assignment mechanism is thought to turn CSs into valuable and wanted stimuli that are attributed with incentive salience and elicit approach and avoidance responses [12–14]. Aspects of this were visible in the impact of CS value on gaze, attention and through pupil dilation also arousal. While outcome-specific PIT may depend in part on model-based mechanisms, the current paradigm has previously been argued to capture mostly outcome-general effects [34]. Hence, in sign-trackers the model-free learning algorithm based on RPEs seems to assign predictive value or incentive salience to the CS, turning it into a motivationally relevant – and wanted – stimulus that elicits responses [12–14] across multiple modalities including in visual attention, arousal, and approach/avoidance behavior.

While all our learning measures indicated model-free learning in sign-trackers, model-based learning in goal-trackers was likewise consistently evident across all measures. As a key property, model-based learning algorithms construct a model of the state transitions, where they estimate the

probabilities for state transitions in a task. Such transition probabilities can be estimated via SPE in the intraparietal sulcus [4], which we found to be stronger in goal-trackers than in sign-trackers. This key novel finding provides evidence for theoretical predictions [2,3] whereby goal-trackers rely more on model-based reinforcement learning. Notice that our conditioning task did not experimentally manipulate state transitions, limiting our access to the process of model construction.

Model-based learning relies on uncertainty to guide selective attention (i.e., associability) [37] and pupil dilation [27], and is thought to be more resistant to Pavlovian response biases compared to model-free control [17,34]. These model-based signatures were present in goal-trackers. Increased associability associated with situations of model-based uncertainty may thus also attract visual attention to the CS to increase its perceptual, but also higher-level processing. These findings converge on the functioning of a model-based system in goal-trackers that learns predictions about upcoming US identity by selective processing of uncertain predictors.

Taken together, these results demonstrate a double dissociation between model-free RPE learning mediating incentive salience attribution in sign-trackers versus model-based SPE learning guiding uncertainty-based selection in goal-trackers. Nevertheless, there are a number of limitations to our findings. First, we used a trace conditioning task with fixed interstimulus interval. The trace conditioning was employed to allow us to examine gaze unconstrained by stimuli being present. The fixed interval and deterministic aspects were employed to ensure the necessary predictability for the eye-tracking analyses. While we judged these design choices to be necessary, they are likely also responsible for the relatively weak nature of the RPE signals observed. Second, the RPE signal was specific for rewards, and was not identifiable for losses [31]. Computational modeling, however, did not show differences between model-free learning responses in either gaze or pupil size in terms of either mechanisms or parameters for rewards versus losses, and defining sign- and goal-trackers based on appetitive trials did not reveal the same distinctions across modalities. Other approaches to define sign- and goal-trackers that were based on all trials, by replacing the gaze index with the

probability to fixate the CS, and by using a computational model of gaze responses rather than a linear regression of CS value showed similar convergent effects across gaze, behavior and MRI measures. The failure to see such distinction when examining rewards only may relate to the reduction in statistical power (already low due to the deterministic trace conditioning paradigm) when removing half the trials with losses. However, it may also hint at hitherto poorly understood distinctions between learning from rewards and losses. The fact that RPE signals in sign-trackers were specific for appetitive win-associated trials is compatible with the known asymmetric encoding of RPEs in dopaminergic signals. Third, the study and the PIT task were part of a larger study examining learning in a population of individuals at risk of developing alcohol dependence, and our sample therefore consists of 18 year old males drawn from the general population. Since we only investigated male subjects, the results can only be generalized to males. Fourth, one limitation of our study is that visual input differs between sign- and goal-trackers (as their definition is based on gaze fixations). This implies that differences in visual input could bias the fMRI results. However, such biases are unlikely because (a) we analyze brain activity in regions known to encode prediction error signals and (b) because we study highly specific computational prediction error signals. Nevertheless, we performed control analyses, controlling for gaze-dependent visual effects in the fMRI analyses. Our results remained stable in these control analyses, suggesting they are not driven by differences in visual input, but due to differences in prediction error signaling.

Being able to measure sign- and goal-trackers in humans may be useful for investigating disorders such as drug addiction. Drug addiction is strongly tied to the dopamine system. This is thought to be sensitized by repeated drug consumption, leading to increasingly stronger incentive salience attribution to drug-predictive cues, which elicit drug craving and consumption [38]. Sign-tracking animals are known to be more prone to develop addiction [2,39]. In humans, drug consumption in alcohol dependent patients is closely linked to PIT and associated neural activation in the NAc [40,41]. Human sign- and goal-tracking have so far not been studied in addiction but given that sign-tracking can be bred true, it suggests a potential link between familial risk, learning and addiction.

Methods

We confirm that our research complies with all relevant ethical regulations. Ethical approval for the study was obtained from the ethics committee of Charité-Universitätsmedizin Berlin (EA1/157/11) and Universitätsklinikum Dresden (EK228072012); procedures were in accordance with the declaration of Helsinki. Informed consent was obtained from all human participants. Participants received a monetary compensation of 10 Euros/hour for study participation plus a performance-dependent compensation. The data were collected as a part of the LeAD study (www.lead-studie.de; clinical trial number: NCT01679145).

Definition of sign- and goal-tracking

To define sign- and goal-trackers, we computed a gaze direction index as the proportion fixation time on the CS minus the proportion fixation time on the US location during the third second of CS presentation, i.e., gaze index = p(CS) – p(US). A gaze index of 1 indicated 100% of fixation time was spent on the CS, a gaze index of -1 indicated 100% of fixation was spent on the US, a gaze index of 0 indicated the same percentage of fixation times on CS and US, and intermediate values indicated intermediate fixation statistics. For each individual subject, we computed a linear regression of gaze index on the true value of the CS (i.e., -2, -1, 0, +1, +2 Euro). Subjects for whom gaze is attracted more to win-predictive than to loss-predictive CSs have a positive regression coefficient of CS value, and we defined the third of subjects with the most positive regression coefficients as sign-trackers (N = 43). Subjects for whom gaze is attracted towards the goal for expected wins more than for expected losses have a negative regression coefficient, and we defined goal-trackers as the third of subjects with the most negative regression coefficient (N = 43).

Overview of computational modeling

Model-based valuation estimates the probabilities for arriving at each outcome state *j* (US) given the observation of a certain cue *i* (CS), which can be written as a matrix of transition probabilities: $T_{i,j} = p(US_j|CS_i)$. Moreover, it estimates a 'reward matrix', which stores expected value V for each US

outcome *j*: $R_j = E\left[V|US_j\right]$. When a CS *i* is presented, the model-based system determines the expected value V_i^{MB} by considering all possible US outcomes, and by weighing their expected values by their transition probabilities. This can be formulated by multiplying the transition matrix with the reward matrix: $V_i^{MB} = T_i \times R$. Learning in the model-based system involves learning the transition matrix from the experience of state prediction errors (SPE), δ_t^{SPE}. Initially, before learning, all five US outcome states in our task are equally likely, i.e., $T_{i,j} = \frac{1}{5} = 0.2$. Subjects experience a state prediction error when they encounter a transition from a CS *i* to a US *j* in trial *t*, $\delta_t^{SPE} = 1 - T_{i,j}^t$. They use this to update the transition matrix, $T_{i,j}^{t+1} = T_{i,j}^t + \eta \times \delta_t^{SPE}$. To keep the matrix normalized to total probabilities of one, transition probabilities for the other US $j' \neq j$ that have not been observed are reduced by $T_{i,j'}^{t+1} = T_{i,j'}^t \times (1 - \eta)$. We assume that the reward matrix is known instantly and with certainty. We approximate uncertainty (*Unc*) in state predictions as $Unc_t = 1 - \max_j T_j^t$. We repeated core analyses with a fully Bayesian model-based learner with explicit uncertainty computation, with comparable results.

Model-free learning, to the contrary, conflates the transition matrix and the reward matrix and thus does not learn about the identity of US outcomes. Instead, it uses the experience of reward prediction errors (RPE), $\delta_t^{RPE} = R^t - V_{i,t}^{MF}$, where R^t is the actually experienced reward value in trial *t* and $V_{i,t}^{MF}$ is the current model-free state value for CS *i* in trial *t*, to update estimates of expected value directly: $V_{i,t+1}^{MF} = V_{i,t}^{MF} + \alpha \times \delta_t^{RPE}$.

We assume that model-based learning influences gaze direction and pupil size based on trial-by-trial state uncertainty (Model *Unc*) and accordingly predict the dependent variable as $\widehat{y_t^{id}} = int_{Unc} + b_{Unc} \times Unc_t$, where $\widehat{y_t^{id}}$ of subject *id* for trial *t* is the predicted gaze direction index during the third second of CS presentation or the predicted pupil size in the last second before US presentation, int_{Unc} is a free intercept or baseline parameter capturing the expected value of the dependent variable after learning and complete state certainty, b_{Unc} is a free parameter for the weight of

model-based uncertainty, which we re-parameterize to $b_{Unc} = e^{b_{Unc}'}$, and which measures the degree to which maximal uncertainty before learning biases the gaze index towards the CS relative to baseline.

For model-free learning, we assume that trial-by-trial CS value (Model *Val*) influences gaze and pupil size via a Pavlovian response bias, and predict the observations $\widehat{y_t^{id}} = int_{Val} + b_{Val} \times V_t^{MF}$, where int_{Val} is a free intercept or baseline parameter capturing the average gaze direction or pupil size, b_{Val} is a free parameter measuring the weight of model-free value influencing gaze or pupil size, i.e., the Pavlovian response bias.

For gaze direction and pupil size, we assume the likelihood for individual observations y_t is normally distributed $p(y_t|\widehat{y_t}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[(y_t - \widehat{y_t})^2/(-2\sigma^2)]$, where y_t is the observed dependent variable per trial, $\widehat{y_t}$ is the prediction by the learning model, and σ^2 is the residual variance. Note that the distribution of the gaze direction index deviates from normality. The likelihood for the gaze direction index $Gaze^{id} = \{y_t^{id}\}_{t=1}^{T}$ per subject *id* across *T* trials, is then $p\left(Gaze^{id}\middle|\widehat{y^{id}}, \sigma_{id}^2\right) = \prod_t p\left(y_t^{id}\middle|\widehat{y_t^{id}}, \sigma_{id}^2\right)$. For the pupil size $Pupil = \left\{\{y_t^{id}\}_{t=1}^{T}\right\}_{id=1}^{N}$ of *N* subjects with each *T* trials, we pool the likelihood across all subjects: $p(Pupil|\hat{y}, \sigma^2) = \prod_{id}\prod_t p\left(y_t^{id}\middle|\widehat{y_t^{id}}, \sigma^2\right)$.

We studied signals of model-free RPE and model-based SPE using fMRI. For the model-free RPE signal, we determined the trial-by-trial temporal difference RPE for CS and US onsets [1,11]. Onset of the CS changes model-free value expectation from 0 (at trial onset) to the predictive value of the CS, $V_{i,t}^{MF}$, yielding a temporal difference RPE of $\delta_{CS}^{RPE} = V_{i,t}^{MF}$. At US onset, value expectation changes from the predictive value of the CS, $V_{i,t}^{MF}$, to the observed US value, R^t, i.e., $\delta_{US}^{RPE} = R^t - V_{i,t}^{MF}$ [1,11]. We combined these two RPE regressors into a single regressor coding model-free temporal difference RPE. Moreover, we took the model-based trial-by-trial SPE signal $\delta_t^{SPE} = 1 - T_{i,j}^t$ to modulate fMRI activity at the time of US onset. The RPE and the SPE regressors were each entered

as a parametric modulator, either at the time of CS and US onset (RPE), or at the time of US onset (SPE). Each modulated onset regressors with onset durations equal to the 3 sec of stimulus presentation.

The following sections provide more detailed information on the used methods.

Paradigm

The paradigm tested Pavlovian-instrumental transfer (PIT) [21,42] and consisted of four parts: (1) instrumental conditioning, (2) Pavlovian conditioning, (3) PIT, and (4) a forced choice task (Supplementary Fig. 1). Instrumental conditioning was conducted before and the forced choice task after the scanning session; Pavlovian conditioning and PIT were assessed during functional magnetic resonance imaging (fMRI). The task was programmed using Matlab 2011 (MATLAB version 7.12.0, 2011; MathWorks, Natick, MA, USA) with the Psychophysics Toolbox Version 3 extension [43,44]. It was presented on a computer screen (instrumental training, forced choice) and on a projector via a mirror system (Pavlovian conditioning, PIT). For a detailed description of the paradigm see [40,45].

Instrumental conditioning

Subjects were instructed to collect or avoid shells by repeated button presses. To collect a shell, subjects had to move a red dot (Supplementary Fig. 1a) onto a shell by repeated button presses, and otherwise did not collect the shell. Each response moved the dot a fraction of the way towards the shell but this was not shown on screen. At least five button presses (two-second response window) were needed to collect a shell, which subjects were not informed about. Subjects received probabilistic feedback. On approach trials, a “good” shell was monetarily rewarded in 80% and punished 20% of trials if collected and vice versa if not collected. On non-approach trials, if a “bad” shell was collected, this was monetarily punished in 80% and rewarded in 20% of the trials, and vice versa if not collected. Participants learned to respond to three “good”, i.e. approach, and three “bad”, i.e. non-approach, shells through trial and error. Participants performed 60-120 instrumental

training trials, depending on their performance: in order to ensure that all subjects were at comparable performance levels before advancing to the PIT part, a learning criterion was enforced (80% correct choices over 16 trials).

Pavlovian conditioning

At the beginning of each trial, a compound CS consisting of fractal-like pictures and pure tones (henceforth referred to as 'fractal CSs') was presented for three seconds. This was followed by a delay of three seconds with two fixation crosses at the two potential CS locations (left and right; Supplementary Fig. 1b). Finally, the US was presented for three seconds at the position opposite to where the CS had been presented.

The set of stimulus pairings consisted of two positive CSs paired with images of +2€ and +1€ coins, one neutral CS paired with 0€, and two negative CSs paired with -1€ and -2€ (coins with a superimposed red cross, see also Supplementary Fig. 1a). The identity of the fractal and the height of the tones deterministically predicted US value such that higher tones predicted higher/lower values, with the mapping counterbalanced across subjects. Moreover, there was an initial shaping period. First all Pavlovian CSs were presented in descending order (+2, +1, 0, -1, -2 €) and then in ascending order.

Subjects were instructed to observe the CSs and USs and to memorize the pairings. All subjects completed 80 trials, in which each of the five different CSs was presented 16 times in a random (except for initial shaping) sequence with randomized (left versus right) stimulus locations.

PIT

Subjects then performed 90 trials of the instrumental task just as during training, but with fractal CSs tiling the background (Supplementary Fig. 1c). No outcomes were presented, but subjects were instructed that their choices still counted toward the monetary outcome. Each of the six

instrumental shells (three shells for each of the two conditions: instrumental approach/non-approach) was presented with each of the five Pavlovian CSs a total of three times (6 x 5 x 3 = 90 trials), such that instrumental and Pavlovian approach/non-approach were orthogonalized. This was implemented to control for instrumental approach and non-approach tendencies during PIT (cf. [17,21]).

There were also interleaved trials with drink-related stimuli (alcoholic drinks, water) tiling the background. Results for these will be reported separately.

Forced-choice task

Finally, subjects chose one of two sequentially presented compound CSs (Supplementary Fig. 1d). They received 10% of the monetary US value associated with the chosen option and were fully instructed about this. Each of the 10 possible CS pairings was presented three times in an interleaved, randomized order, yielding a total of 30 trials. Within a trial, CSs were presented one at a time for two seconds each. Slow responses led to a reminder requesting faster responses.

Participants

The two-centre study was conducted in Berlin and Dresden, Germany. We assessed 198 participants; all male and all with the same age of 18 years. The data were collected as a part of the LeAD study (www.lead-studie.de; clinical trial number: NCT01679145).

Exclusion criteria were left-handedness, a history of any substance dependence or current substance use (assessed by breath and drug urine testing) except for nicotine dependence, other major psychiatric disorders (DSM-IV axis I; CIDI; [46]) and neurologic disorders.

Here, we report on a subsample of 144 subjects for which eye-tracking data was available during Pavlovian conditioning. These same subjects were tested on all tasks, i.e., we tested the same sample repeatedly. For each task and recording technique data for some subjects was missing for technical reasons (e.g., recording failures or early task abortion); for numbers of missings see below.

The study was designed and sample size of 198 subjects was chosen to detect moderate differences (a priori Power analysis: $d = 0.4, \alpha = 0.05, \beta = 0.80$) in learning parameters or brain activity in healthy adults with high versus low risk alcohol consumption, using longitudinal follow-up over three years. For our cross-sectional analysis, valid eye-tracking data in the fMRI scanner was available for 129 subjects. The statistical group identification (described in the manuscript) resulted in a final sample size of 43 subjects per sign-/goal-tracker group, which yielded a good power to detect medium differences in learning parameters or brain activity ($d = 0.6, \alpha = 0.05, \beta = 0.79$).

Randomization

There were no experimental group allocations. Group definitions in the analyses were based on statistical tests described in the methods. Assignment of experimental stimuli (Pavlovian CSs; instrumental shells) to experimental (reinforcement) conditions, stimulus orderings and locations across trials were randomized.

Blinding

There were no experimental group allocations. Group definitions in the analyses were based on statistical tests described in the manuscript.

Measurements

Eye-tracking

We recorded eye-position and pupil size during Pavlovian conditioning via an EyeLink 1000 eye-tracker (SR Research; recording binocularly at 1000 Hz; in Dresden) and via an iViewX MRI-LR eye-tracker (SMI; recording monocularly at 50 Hz; in Berlin), which were both used in the fMRI scanner via a mirror system mounted on the head coil. Calibration of the eye-tracker was performed inside the scanner before the start and after 40 trials of Pavlovian conditioning. At the beginning of each trial subjects were instructed to fixate a central fixation point. Failures to fixate lead to a reminder (a maximum of two times per calibration).

fMRI acquisition

Functional imaging was performed on two Siemens Trio 3 Tesla MRI scanners with echo planar imaging (EPI) sequences (repetition time: 2410 ms; echo time: 25 ms; flip angle: 80°; field of view: 192 x 192 mm^2; voxel size: 3 x 3 x 2 mm^3) comprising 42 slices approximately -25° to the bicommissural plane. For coregistration and normalization during preprocessing, a three-dimensional magnetization-prepared rapid gradient echo image was acquired (repetition time: 1900 ms; echo time: 5.25 ms; flip angle: 9°; field of view: 256 x 256 mm^2; 192 sagittal slices; voxel size: 1 x 1 x 1 mm^3). Prior to functional scanning, a field map was collected to account for individual homogeneity differences of the magnetic field.

Participants wore MR-compatible Siemens head-phones. Responses were made on a 1 x 4 current design MR-compatible response box button using the dominant index finger (instrumental response in training and transfer) or two buttons using the left and the right index finger (forced choice).

Data analyses and statistics

Data were analyzed using Matlab 2013a (MATLAB version 8.1.0.604, 2013; MathWorks, Natick, MA, USA) and the R System for Statistical Computing [47]. fMRI data were analyzed using Statistical Parametric Mapping 8 (SPM8; Wellcome Department of Imaging Neuroscience; http://www.fil.ion.ucl.ac.uk/spm/).

For eye-tracking analyses we performed repeated measures ANOVA using the R-package *afex* [48]. Contrasts were computed using the R-package *emmeans* [49]. For fMRI analyses, at the second level we performed either one-sample Student t-tests or two-sample Welch t-tests (capturing situations of equal and of unequal variances) [50]. For random effects analyses of behavioral responses, we performed Shapiro-Wilk tests to test the normal distribution assumption of t-tests. If violated, non-parametric bootstrapping with 1,000,000 case resamples and bias-corrected adjusted (BCa) confidence intervals (90%, 95%, 99%, and 99.9%; R-package *boot* [51,52]) was used instead. Statistical tests were two-tailed unless otherwise noted. Error bars in Figures represent repeated measures S.E.M. [53]. Error bars for pupil analyses (Fig. 2e+f) are extracted via linear mixed effects models. Box-and-whisker plots show the following statistics: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. For F-tests, as a measure of effect size we report the proportion of variance of the dependent variable accounted for by the levels of the factor (i.e., n_p^2), together with 95% confidence intervals (CI), as computed by the function ci.pvaf() from the R-package MBESS.

Eye-tracking analyses - gaze

Preprocessing

Data from the EyeLink 1000 system (right eye) were down-sampled to the 50 Hz that was available for the iViewX system. Given different sampling rates between testing sites we checked for

differences in gaze index results (Fig. 1b-d, Supplementary Fig. 2a) but found no significant difference ($p > .1$). We corrected for temporal-spatial drifts and distortions of the eye-tracking data for each subject and each calibration across trials. 15 subjects showed poor correction performance and were removed from analysis after visual inspection, yielding 129 subjects with valid eye-tracking data. We repeated some core analyses using uncorrected eye-tracking data, and found overall consistent results.

No valid gaze data was recorded after the second calibration in three out of the 129 valid subjects (1 sign-tracker, ST, 2 Controls). In one (ST) subject, no eye-tracking data was available for the last 22 trials. Additionally, an average of 12.8% of the eye-tracking samples recorded during CS presentation (Median = 10.6%, SD across subjects = 10.0%) was missing or invalid, including gaze samples outside screen boundaries or during blinks as detected by the eye-tracker device. This yielded an average of 0.8% (Median = 0.0%, SD = 2.2%) trials with no valid gaze data overall, and for the third second of CS presentation an average (SD) of 3.5% (6.6%) trials with no valid gaze data.

Valid gaze-samples were classified as being directed at one of three spatial regions of interest (ROIs): (*i*) the CS, (*ii*) the spatial location of later US presentation, and (*iii*) the rest of the screen reflecting the background. Note that for each second of CS presentation, the percentage of samples within each ROI (pROI) also reflects the cumulative gaze times, i.e., dwell times, for these ROIs, proportionally corrected for missing/invalid data.

Percentage fixation times

CS onset is known to trigger initial orienting responses [22,54] that do not differ between ST and GT [24]. Later on, gaze exhibits Pavlovian conditioned responses to CS value, visible in enhanced fixations on appetitive compared to neutral or aversive cues [22,54–56] (for early responses see [57]). Moreover, uncertainty is known to attract attention to the CS [37], which decreases across learning. To identify

Pavlovian conditioned responses in gaze, we estimated how the influence of Pavlovian (CS) value (+2, +1, 0, -1, -2 €) on percentage fixation times on the CS, the US-location, and the background differed between the three seconds of CS presentation via repeated measures ANOVA. Moreover, we tested how attention changed across learning by performing random effects linear regression analyses, regressing percentage fixation times on trial number (mean-centered). Regression coefficients were analyzed via repeated measures ANOVA to test how trial effects differed between the CS, the US location, and the background (factor location), between CS value levels (+2, +1, 0, -1, -2 €), and between the three seconds of CS presentation. We followed up on significant interactions using post-hoc contrasts. Moreover, based on the results from these analyses we tested the interaction CS value x trial number x location for the third second of CS presentation. We used planned contrasts to test linear CS value effects (-2, -1, 0, +1, +2 €). Last, we tested a contrast coding whether an increase in CS value effects across trials on the CS was stronger than that on the US.

Gaze index

To assess sign- and goal-tracking, we computed a gaze index measuring the difference in the probabilities of approaching (here fixating) the CS minus the US-location. The aim was to parallel the approach employed in animal research to measure relative approach to a CS and US [10]. The gaze index was 1 if the entire time was spent on the CS; -1 if the entire time was spent looking at the US; 0 if there was no preference for either CS or US-location; and it had some intermediate value if gaze was distributed between the CS, the US location, or the background. For example, for values of p(CS) = 0.7, p(US) = 0.2, and p(BG) = 0.1, the gaze index would be p(CS) - p(US) = 0.7 - 0.2 = 0.5.

We investigated learning by testing whether the gaze index decreased across trials, whether it increased with increasing CS value, and whether the observed CS value effect got stronger across trials (interaction CS value x trial number per stimulus; [19,22,37,55]).

Definition of ST and GT based on gaze index

To define ST versus GT we computed the influence of CS value on the gaze index during the third (i.e., last) second of CS presentation per subject. ST were defined as the third with the most positive regression coefficients (N = 43) and GT as the third with the most negative (N = 43). We tested whether the frequency of ST versus GT differed between testing sites (Berlin / Dresden) using a chi-squared test.

We tested effects of CS value on the gaze index, and whether CS value effects got stronger across trials (one-tailed test for ST [19]). Moreover, for sign-trackers we separately tested the effect of trial number for trials involving wins versus losses. Fig. 1d+f visualize how influences of CS value developed over time in ST versus GT.

Model-based influences on gaze

Learning in the model-based system involves learning the transition between CSs and USs. Cues for which predictions were more uncertain were hypothesized to be attended more to support optimal processing and learning [37]. Specifically, we formulated a state transition matrix T(CS,US) of transition probabilities, where each element in the matrix holds the current estimate for the probability of transitioning from state CS to US. At the beginning of learning, the probability to observe one of the five different outcomes after seeing a certain CS is T_0(CS,US) = 1/5 as all USs are equally likely. In each conditioning trial t, the model-based system computes a state prediction error (SPE):

$$\delta_t^{SPE} = 1 - T_t(CS, US) \quad (1)$$

and updates the probability T(CS,US) of the observed transition via:

$$T_{t+1}(CS, US) = T_t(CS, US) + \eta \cdot \delta_t^{SPE} \quad (2)$$

where the free parameter η is a learning rate. For the other USs not observed in the current trial (i.e., all US' except for the observed US), the estimated probabilities are updated to keep probability distributions normalized using the equation $T_{t+1}(CS, US') = T_t(CS, US') \cdot (1 - \eta)$.

We approximate uncertainty (U) in state predictions in this experience-based model-based system via the distance of the highest state transition probability conditional on the visited CS "cs" from certainty, that is $U = 1 - \max[T(CS = cs, US)]$. We assumed that predictive uncertainty directly increases attention towards the predictive CS, and hence increase CS-related eye fixations. We therefore modeled the trial-by-trial gaze index via:

Model MB: $$\mathrm{GazeIndex}_t = c + \beta_U^{gaze} \cdot U_t(s) \ , \quad (3)$$

where the $\mathrm{GazeIndex}_t$ was computed during the third second of CS presentation in trial t, c is a free constant baseline parameter capturing preference for CS- over US-related fixations after learning and complete state certainty (e.g. capturing effects of visual salience [58]), $U_t(s)$ is the trial-by-trial CS-related uncertainty, and β_U^{gaze} is a free parameter for the degree to which maximal uncertainty during the first trial biases the gaze index towards the CS relative to baseline. Here, we multiplied the β_U^{gaze} parameter by the maximal uncertainty in the first trial, i.e., 0.8, to standardize the uncertainty-based weight to reflect the effect of maximal uncertainty in our experimental design.

Model-free influences on gaze

In model-free learning, the value of CSs was learned from experience via errors in predicting the US outcome value. A simple model-free reinforcement learning (RL) model computes a reward prediction error (RPE):

$$\delta_t^{RPE} = R_t - V_t(s) \ , \quad (4)$$

and updates the expected CS value $V_t(s)$ via:

$$V_{t+1}(s) = V_t(s) + \alpha \cdot \delta_t^{RPE} \quad , \tag{5}$$

where $V_t(s)$ is the value of a certain CS s in trial t, R_t is the value of the US, and α is a free learning rate parameter. We assumed that the trial-by-trial value estimate $V_t(s)$ exerts a Pavlovian response bias on gaze direction, and we hence modeled the influence of model-free learning on trial-by-trial gaze index in the model "Value" via:

Model MF: $$\mathrm{GazeIndex}_t = c + \beta_V^{gaze} \cdot V_t(s) \ , \tag{6}$$

where β_V^{gaze} is a free parameter controlling the weight of the model-free Pavlovian response bias from CS value. Positive values of the β_V^{gaze} weight parameter indicate a sign-tracking response whereas negative weight values indicate goal-tracking.

Dual model-free and model-based influences on gaze

Last, we constructed a model assuming dual learning systems for model-free value ($\omega = 0$) and for model-based uncertainty ($\omega = 1$) are combined via a weighting parameter ω to guide attention:

Model MF + MB: $$\mathrm{GazeIndex}_t = c + \beta^{gaze} \cdot \Big([1 - \omega] \cdot V_t(s) + \omega \cdot \widetilde{U}_t(s) \Big) \ . \tag{7}$$

In the present task, uncertainty was constrained between 0.8 and 0, whereas CS value ranged between -2 and +2. Direct comparison of uncertainty and CS value is therefore difficult. For a comparison via the weighting parameter, we therefore normalized the uncertainty variable to span the same range as CS value by computing $\widetilde{U}_t(s) = (U_t(s) - 0.4) \cdot \frac{2}{0.4}$. Note, that it is therefore

difficult to interpret the absolute size of the weighting parameter ω, but that it is useful to analyze differences in parameter estimates between groups.

Parameter estimation and model comparison

To perform model comparison, we estimated free model parameters using maximum likelihood estimation (MLE) for each individual subject, assuming Gaussian residuals. Bounded parameters were transformed to an unbounded scale for fitting: learning rate parameters for learning from reward prediction errors or state prediction errors as well as the weighting parameter ω were bound to values between 0 and 1 via the logistic transform $\alpha = \frac{1}{1+\exp(-a)}$; the uncertainty-based weight b_U was bound to positive values via an exponential transform $\beta_U = 0.8 \cdot \exp(b_U)$. Optimization was performed using the *nlm* function in the R-package *stats* [47]. We compared models for each subject by computing the difference in BIC values. We tested whether BIC values differed between sign- and goal-trackers via repeated measures ANOVA. We used contrasts to test our hypotheses of (i) a stronger value effect in sign-trackers, i.e., stronger evidence for the model assuming conditioned responses to CS value [13,16,17], and (ii) a stronger model-based [2,3] uncertainty response in goal-trackers.

To increase stability in the estimation of noisy model parameters, we followed up MLE via fixed effects maximum a posteriori (MAP) estimation. We assumed weakly informative independent Gaussian priors with mean zero (except for the learning rate parameters, where we assumed a mean prior learning rate of $\mu_\alpha = 0.3$, i.e., $\mu_a = \log(0.3/0.7)$), and a standard deviation of $\nu = 5$.

Eye-tracking analyses of pupil dilation

Pupil size data for valid fixation samples was z-standardized for each subject. For each subject and calibration (trials 1-8 and trials 9-16) we corrected for average baseline pupil size during one second

before CS presentation. We removed data from the first two trials per CS to prevent potential biases arising from the fixed order of stimulus presentation in these trials. We analyzed pupil size via repeated measures ANOVA with factors trials (3-8 vs. 9-16), CS value (-2 to +2 €), and time within trial (six seconds from CS onset to US onset). We hypothesized that pupil size during the last second of US anticipation should decrease from the beginning to the end of conditioning, reflecting decreasing uncertainty with learning [27]. Moreover, we expected pupil size to increase for expected wins compared to expected losses or neutral outcomes [28], and we coded planned contrasts for linear CS value effects. This CS value effect should increase across trials and we tested interactions of linear CS value with trials (3-8 vs. 9-16). To minimize influences from luminance, we tested effects nested within the last second before US presentation. We tested whether effects of CS value and of trials differed between sign- and goal-trackers. Contrasts tested effects separately for sign- versus goal-trackers and for trials 3-9 versus 9-16. In a first overall approach, we studied effects in all six seconds from CS onset to US onset. We thus tested whether effects of CS value and of trials changed as a function of linear time within trials (seconds 1-6 after CS presentation; i.e., interaction trials x time, and interaction CS value x trials x time). Next, we focussed analysis on the last second before US onset, where the signal is least confounded by luminance-related influences from CS presentation. Estimated contrasts and standard errors were extracted from the ANOVA for visualization.

To visualize effects of trials and of CS value, we moreover performed random effect linear regression analyses on data before US onset, regressing pupil size on CS value separately for trials 3-8 versus trials 9-16 for time bins of 100 ms each. For each time bin and experimental half, we excluded outlier subjects with CS value effects deviating more than six standard deviations from the mean, yielding a total of two excluded data points. We performed repeated measures ANOVAs on the estimated regression coefficients for (a) the intercept and (b) the linear CS value effect, with factors time bin, trials (3-8 versus 9-16), and group (sign- versus goal-trackers). We then performed exploratory tests nested within each time bin (a) of the trial effect (3-8 versus 9-16) additionally nested within sign-

versus goal-trackers, and (b) of the difference between sign- and goal-trackers in the CS value effect, nested within trials (3-8 versus 9-16).

Pupil analyses: Computational modeling

We fitted computational learning models to the trial-by-trial pupil data. To this end, we extracted average pupil size per trial for the last second before US presentation, where influences from luminance should be minimal. We tested two different computational models.

First, we used the model-free reinforcement learning model (see Equations (4) and (5)) to obtain the trial-by-trial value of the CS, $V_t(s)$, which was assumed to modulate pupil size via a weight parameter β_V^{pupil} via:

$$\text{pupil}_t = c + \beta_V^{pupil} \cdot V_t(s) \ . \quad (8)$$

Second, we used the model for model-based state learning (see Equations (1) and (2)) to obtain the trial-by-trial state uncertainty, $U_t(s)$, which was assumed to modulate pupil size via a weight parameter β_U^{pupil} via:

$$\text{pupil}_t = c + \beta_U^{pupil} \cdot U_t(s) \ . \quad (9)$$

Again, c is a constant, here capturing pupil size independent of learning. As for modeling of gaze direction, we again performed maximum a posteriori (MAP) estimation of the learning rate parameter α, the regression coefficient parameter β^{pupil}, and the residual variance σ. For parameter estimation, the learning rate parameter was again transformed to a bounded scale between 0 and 1 with the sigmoid transform $\alpha = 1/(1 + \exp(-a))$. Moreover, we constrained the uncertainty-based weight to theoretically expected positive values using an exponential transform

$\beta_U^{pupil} = \exp(b_U^{pupil})$. We used weakly informative Gaussian priors with a prior mean for the learning rate of $\mu_\alpha = 0.3$ (i.e., $\mu_\alpha = \log(0.3/0.7)$), a prior mean for the regression parameter of $\mu_\beta, = 0$, and prior standard deviations of $\nu = 5$. Due to the large noise in pupil size, we obtained fixed effect maximum a posteriori (MAP) estimates for sign-trackers and goal-trackers via Newton-type minimization with the function *nlm* from the *stats*-package in the R System for Statistical Computing. To test both models against each other, we computed BIC values, and computed the difference in BIC between models for sign- and goal-trackers separately.

For visualization of trial-by-trial effects of CS value, we aimed to maximize sensitivity within trials. To this end, we removed between-trial variance in the intercept by subtracting the average pupil size per trial and per CS. Moreover, to normalize CS value effects and remove trends in average pupil size across trials we performed z-transformation across the five different levels of CS value for each trial separately. Linear mixed effects models were used to estimate the effect of CS value in sign- and in goal-trackers for each trial. Moreover, the computational value model was re-fitted to this normalized data for visualizing model predictions. The results from these analyses are shown in Figure 2f.

Behavioral analyses

Forced choice task

Data for the forced choice task was available for 39 GT subjects and for 42 ST subjects. Successful Pavlovian learning was assessed via the percentage of correct choices in the forced choice task. We tested for a group-difference in percentage correct choices.

Instrumental conditioning

Data on instrumental conditioning was available for all 43 ST and 43 GT. We measured overall learning speed as the number of trials needed until reaching the learning criterion (with a minimum of 60 and a maximum of 120 learning trials), and tested learning speed in ST versus GT. Moreover, to measure initial learning we extracted the first twenty trials. To measure asymptotic learning, we extracted the last twenty trials. For these, we computed the difference in the response rates between instrumental conditions (collect versus leave) for each subject, and tested the effect in ST versus GT.

PIT

Data on the PIT task was available for all 43 ST, and for 41 GT. We calculated individual PIT effects by regressing the number of button presses on the five different Pavlovian values, and tested whether PIT effects were larger than zero for individual subjects via t-tests. We tested the strength of the PIT effect in ST and in GT, and performed one-tailed tests of the a priori hypothesis [19] that PIT effects are stronger and more frequently individually significant in ST compared to GT.

fMRI analyses

Preprocessing

fMRI recordings were preprocessed using Nipype [59]. First, correction for differences in slice time acquisition to the middle slice was performed. Voxel-displacement maps were estimated based on acquired field maps. All images were realigned to correct for head motion, distortion and their interaction. After co-registration of the individual structural T1 images to the individual mean EPI, the structural image was spatially normalized with a resampling resolution of 2 x 2 x 2 mm^3 and the normalization parameters were applied to all EPI images. Finally, images were spatially smoothed with a Gaussian kernel of 8 mm full width at half maximum. Prior to statistical analysis, data were high-pass filtered with a cut-off of 128 sec.

Pavlovian conditioning: Value learning

We performed model-based fMRI analyses via 1st- and 2nd-level analyses in SPM. We used the model-free reinforcement learning (RL) model (see Equations (4) and (5)) to compute the trial-by-trial value of the CS V_t. Based on the RL model, we determined the trial-by-trial temporal difference reward prediction error (RPE) for CS and US onsets. Onset of the CS changes value expectation from zero (at trial onset) to the predictive value of the CS, $V_t(s)$, yielding a temporal difference RPE of $RPE_{CS} = V_t(s) - 0$. At US onset, value expectation changes from the predictive value of the CS, $V_t(s)$, to the observed US value, R_t, i.e., $RPE_{US} = R_t - V_t(s)$.

The learning rate parameter α was set to 0.05 based on an exploratory analysis in a related sample with the same task setup (unpublished data). This value maximized the signal strength in the NAc. Repeating these analyses with the current sample confirmed the same pattern for the learning rate (section "Varying model-free learning rates" below), but also indicated good robustness wrt. the precise choice. The small value also corresponded to parameter estimates obtained from the pupil size data, which for the sign-trackers yielded a learning rate of $\alpha = 0.06$.

In the first-level SPM model, we included the onsets of CSs as well as USs with their stimulus durations of three seconds within one onset regressor. Stimulus onsets were parametrically modulated by the trial-by-trial temporal difference reward PE. Additional nuisance regressors captured variance specific to US onsets, the calibration after trial 40, fixation reminders, and realignment parameters with derivatives [60]. Regressors were convolved with the canonical haemodynamic response function (HRF).

Animal results suggest a fixed timing and duration of the midbrain dopamine responses [10]. We therefore focused analysis on the main RPE regressor, but controlled for possible individual variance in the onset and duration of the blood oxygenation level dependent (BOLD) response in the current

paradigm, by including temporal and dispersion derivatives of the HRF as nuisance regressors. Control analyses confirmed that there were no significant differences in the delay or the duration of the RPE response in the NAc between ST and GT groups.

Individual subjects' parameter estimates for the reward PE parametric modulator were taken to the second level. Valid fMRI recordings during Pavlovian conditioning were available for 39 ST and 39 GT. A two-sample t-test was performed comparing the RPE effect between ST and GT, with testing site as a control covariate of no interest. Differences between ST and GT in BOLD responses were tested via an F-test. The RPE signal in sign- and goal-trackers was tested via nested contrasts with 78 (n subjects) – 3 (parameters used for the mean signals in ST and GT and for the covariate site) = 75 degrees of freedom. For visualization (Fig. 4a), we computed a contrast coding the a priori hypothesis [10] of a stronger reward PE response in ST compared to GT. Visualization threshold was $p_{uncorrected} < .005, k = 0.$ Statistical testing was performed in an a priori defined volume of interest (VOI) in the bilateral nucleus accumbens (NAc) [10]: we chose a previously validated bilateral ventral striatal VOI a priori from the IBASPM 71 atlas; we derived this from the Wake Forest University (WFU) PickAtlas software (www.fmri.wfubmc.edu/software/PickAtlas). Reward prediction errors in learning tasks akin to ours have been reported in this very VOI on numerous previous occasions (e.g., [61–64] and many others). In addition, this a priori VOI overlaps substantially with a VOI shown in a published meta-analysis to exhibit strong RPE signals [65]: 78% of our a priori VOI were inside the VOI from the meta-analysis. Moreover, we performed a meta-analysis at neurosynth.org of the term “prediction error”. This showed significant prediction-error related activity in 82% of our a priori VOI. Hence it appears beyond doubt given the current state of the scientific literature that our a priori VOI can be validly used to test for RPE signals. We used family-wise error (FWE) correction within the VOI to control for multiple comparisons.

Analyses of appetitive trials

While a wealth of evidence supports positively coded appetitive reward prediction errors in striatal dopamine activity, the coding of aversive reward prediction errors remains less clear. Some evidence suggests aversive RPE may be coded inversely, that is, as a signed prediction error or salience signal [31]. To exclude potential confounds or noise from aversive RPE signals, we repeated the RPE analysis focusing only on win-predictive and neutral CSs (0, +1, +2 €). We coded the onsets of win-predictive and neutral CSs in one onset regressor, while the onsets of loss-predictive CSs were modeled as a separate onset regressor. The win- and neutral-predictive CSs were parametrically modulated by trial-by-trial temporal difference appetitive reward prediction errors. An additional control regressor modulated the loss-predictive onset regressor parametrically by the prediction errors for loss trials. Analyses focused on the prediction error modulator for trials involving wins and neutral outcomes. We extracted the average of the RPE response from the bilateral NAc VOI and performed one-tailed t-tests for a positive RPE signal in each group, and a one-tailed t-test of the a priori hypothesis [10] that the RPE learning signal was stronger in sign- than in goal-trackers.

Prediction-error correlates outside the ventral striatum

Prediction error-like signals are also observed in other regions of the brain reward system, and whether these signals are selectively present in sign-trackers is unknown. Dopaminergic neurons in the ventral tegmental area (VTA [1]) are known to project not only to the NAc, but also to dorsal striatum (Putamen, Caudate), Amygdala, and ventromedial prefrontal cortex (vmPFC) and may drive fMRI BOLD correlates of RPE signals in these areas [66]. Sign-trackers do show increased CS-related activity in a range of different regions of the brain reward system [32], but whether these resemble reward prediction errors, is unclear. Here, we tested for RPE-like signals in several a priori volumes of interest (VOIs) thought to carry RPE-like BOLD responses, including the Putamen, Caudate, VTA, Amygdala and vmPFC. VOIs were taken from [67]. Results are reported for the average RPE signal in these VOIs. We first perform a priori tests using ANOVA with factors group (ST/GT) and VOI. We do so for our a prior analysis involving gains and losses, and in addition for the analysis of gains only.

Moreover, we perform exploratory tests for each group of sign- and goal-trackers (one-tailed test of a signal larger than zero), and we perform one-tailed statistical tests of the hypothesis [32] that the RPE-like signals are stronger in sign- than goal-trackers, which we also correct for multiple exploratory tests. Moreover, we visualize results from voxel-vise analyses based on an uncorrected thresholds of $p_{unc.} < .005, k = 40$, $p_{unc.} < .01, k = 40$ (all VOIs), and $p_{unc.} < .05, k = 40$ (VTA).

Explorative analyses moreover tested for prediction error-like signals at a whole brain level. We performed voxel-based analysis with FWE correction, as well as cluster-based analysis with clusters defined based on a threshold of $p < .005$.

Pavlovian conditioning: State learning

The learning of model-based state transitions relies on state prediction errors (SPE; see section "Model-based influences on gaze"), which are known to be coded in the intraparietal sulcus (IPS) and in the lateral prefrontal cortex (latPFC) [4]. To estimate a neural SPE signal, we used the trial-by-trial state prediction error (see Equation (1)) as a predictor in the fMRI analyses. We adapted the first-level SPM model reported above by removing the RPE from the model, and instead including a parametric modulator with the trial-by-trial mean-centred SPE at the US onset time. Parameter estimates for the SPE regressor were examined at the second level. First, we tested whether SPE predicted BOLD responses for sign- and goal-trackers combined in the IPS and the latPFC via voxel-vise analysis with FWE correction in the *a priori* VOIs, and by extracting the average signal for each VOI. The IPS VOI was obtained by summation of hIP1, hIP2 and hIP3 [68] from the probabilistic brain atlas (Jülich-Düsseldorf cytoarchitectonic atlas) using the Anatomy Toolbox [69]. The lateral PFC VOI was extracted from the WFU PickAtlas software. Based on our *a priori* hypothesis of stronger model-based control in goal- than sign-trackers [2,3] we tested whether the SPE signal was stronger in goal- than in sign-trackers and whether there was an interaction of group x VOI using repeated measures ANOVA on the extracted mean signal per VOI. We followed up on a significant interaction using one-

tailed [2,3] random effects two-sample Welch's t-tests. Moreover, we visualize voxel-based results for the group-difference based on an uncorrected thresholds of $p_{unc.} < .005, k = 40$ and $p_{unc.} < .01, k = 40$.

Alternative classification of sign- and goal-trackers

Using computational modeling to define sign- and goal-trackers

To obtain a second, computational, definition of sign- and goal-trackers, we constructed a computational model assuming that uncertainty and a Pavlovian model-free conditioned response bias would add up to direct attention (model Unc + Value):

$$\text{GazeIndex}_t = c + \beta_U^{gaze} \cdot U_t + \beta_V^{gaze} \cdot V_t(s_t) \ . \quad (10)$$

Note that this is effectively the same model as equation (7). It is parametrized differently in terms of two weights β rather than a trade-off parameter ω to allow a more direct measure of model-free and model-based contributions to gaze control. We estimated model parameters for this model for each individual subject. As before, we performed maximum a posteriori (MAP) estimation, using weakly informative independent Gaussian priors with prior means of zero ($\mu = 0$; except for the learning rate parameters, for which we assumed a prior mean of $\mu_\alpha = 0.3$, i.e., $\mu_a = \log(0.3/0.7)$) and standard deviations of $\nu = 5$. Based on the estimated parameters, we used the weight of the model-free Pavlovian conditioned response bias β_V^{gaze} per subject to classify individuals as sign- or goal-trackers. The third of subjects (n = 43) with the most positive weight parameter were classified as sign-trackers, whereas the third of subjects (n = 43) with the most negative weight parameter were classified as goal-tackers.

We repeated some key analyses with this computational definition of sign- and goal-trackers to test the stability of our findings. Specifically, we tested the hypothesis that model-based uncertainty guides gaze more strongly in goal- than sign-trackers by testing whether the weight parameter of model-based uncertainty on gaze direction β_U^{gaze} was larger in goal-trackers than in sign-trackers via a two-sample Welch's t-test. Moreover, we repeated the analyses testing for a larger PIT effect in sign-trackers. For the neural analyses, we tested whether sign-trackers showed a RPE signal averaged across all tested VOIs and whether it was stronger than in goal-trackers. Likewise, for the neural state prediction error, we tested whether the difference between sign- and goal-trackers differed between VOIs (IPS and lateral PFC), whether the SPE signal was stronger in sign- than goal-trackers in IPS, and whether each group showed a SPE signal different from zero. We used one-tailed tests based on the hypothesis of stronger model-free control in sign-trackers and stronger model-based control in goal-trackers [2,3].

Bayesian model of model-based learning and uncertainty

In our model for model-based learning, we used a simple approximation as a measure of uncertainty. We repeated these simple analyses with a slightly more complex Bayesian model of model-based learning, which computes uncertainty explicitly for each single trial. In this model, given a certain CS i has been presented in trial t, we use a Dirichlet distribution to model the probabilities T_i^t for transitioning to one of the j = 1, .., J possible outcome states: $p\left(T_i^t \middle| CS_i^t, \beta_i^t\right) = \frac{1}{B(\beta_i^t)} \prod_{j=1}^{J} \left(T_{i,j}^t\right)^{\beta_{i,j}^t - 1}$, where evidence for each US j given CS i is $\beta_{i,j}^t = \gamma_{i,j}^t + \eta$, where $\gamma_{i,j}^t$ is the number of observed transitions from CS i to US j throughout the experiment up to trial t, and η are the number of prior observations. In this model of state learning, we computed trial-by-trial uncertainty as the variance of the most likely outcome in the Dirichlet distribution. Trial-by-trial prediction errors were computed by taking one minus the expected value of the observed outcome, $\delta_t^{SPE} = 1 - E\left[T_{i,j}^t\right]$.

Data availability

Data sharing will be based on a) the acceptance by the study team that a valid and timely scientific question, based on a written protocol, has been posed by those seeking to access the data; b) that the role of the original study team will be fully acknowledged. Please contact the corresponding author via email to request access to the data. Safeguarding of ethical standards will be ensured by submission of a study amendment to the Charité and Dresden ethics committees. Data access for questions of scientific integrity may additionally be regulated via the funder.
Figures 1 to 6 and Supplementary Figures 2 to 12 have associated source data.

Code availability

The corresponding author can be contacted with requests for the code. Experimental code is freely available upon request to the corresponding author. Analysis code will be provided with data access.

References

1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
2. Huys, Q. J. M., Tobler, P. N., Hasler, G. & Flagel, S. B. The role of learning-related dopamine signals in addiction vulnerability. *Prog. Brain Res.* **211**, 31–77 (2014).
3. Lesaint, F., Sigaud, O., Flagel, S. B., Robinson, T. E. & Khamassi, M. Modelling individual differences in the form of Pavlovian conditioned approach responses: a dual learning systems approach with factored representations. *PLoS Comput. Biol.* **10**, e1003466 (2014).
4. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).

5. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
6. Dickinson, A. & Balleine, B. The role of learning in the operation of motivational systems. in *Steven's handbook of experimental psychology: Learning, motivation and emotion* **3**, 497–534 (2002).
7. Doya, K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw.* **12**, 961–974 (1999).
8. Friedel, E. *et al.* Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* **8**, (2014).
9. Ernst, M. & Paulus, M. P. Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biol. Psychiatry* **58**, 597–604 (2005).
10. Flagel, S. B. *et al.* A selective role for dopamine in stimulus–reward learning. *Nature* **469**, 53–57 (2011).
11. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
12. Berridge, K. C. & Robinson, T. E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Rev.* **28**, 309–369 (1998).
13. Berridge, K. C. & Robinson, T. E. Parsing reward. *Trends Neurosci.* **26**, 507–513 (2003).
14. Hickey, C. & Peelen, M. V. Neural mechanisms of incentive salience in naturalistic human vision. *Neuron* **85**, 512–518 (2015).
15. Robinson, T. E. & Flagel, S. B. Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. *Biol. Psychiatry* **65**, 869–73 (2009).
16. McClure, S. M., Daw, N. D. & Montague, P. R. A computational substrate for incentive salience. *Trends Neurosci.* **26**, 423–428 (2003).

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

18. Dayan, P. & Berridge, K. C. Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cogn. Affect. Behav. Neurosci.* **14**, 473–492 (2014).

19. Garofalo, S. & di Pellegrino, G. Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. *Front. Behav. Neurosci.* **9**, 163 (2015).

20. Morrison, S. E., Bamkole, M. A. & Nicola, S. M. Sign-tracking, but not goal-tracking, is resistant to outcome devaluation. *Front. Neurosci.* **9**, 468 (2015).

21. Huys, Q. J. M. *et al.* Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput. Biol.* **7**, e1002028 (2011).

22. Gottlieb, J. Attention, learning, and the value of information. *Neuron* **76**, 281–295 (2012).

23. Leclerc, R. & Reberg, D. Sign-tracking in aversive conditioning. *Learn. Motiv.* **11**, 302–317 (1980).

24. Yager, L. M., Pitchers, K. K., Flagel, S. B. & Robinson, T. E. Individual variation in the motivational and neurobiological effects of an opioid cue. *Neuropsychopharmacology* **40**, 1269–1277 (2015).

25. Gottlieb, J., Oudeyer, P. Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* **17**, 585–593 (2013).

26. Renninger, L. W., Verghese, P. & Coughlan, J. Where to look next? Eye movements reduce local uncertainty. *J. Vis.* **7**, 6 (2007).

27. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046 (2012).

28. Manohar, S. G. & Husain, M. Reduced pupillary reward sensitivity in Parkinson's disease. *NPJ Park. Dis.* **1**, 15026 (2015).

29. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl.)* **191**, 391–431 (2007).

30. Rutledge, R. B., Dean, M., Caplin, A. & Glimcher, P. W. Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* **30**, 13525–13536 (2010).

31. Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. Differential encoding of losses and gains in the human striatum. *J. Neurosci.* **27**, 4826–4831 (2007).

32. Flagel, S. B. *et al.* A food predictive cue must be attributed with incentive salience for it to induce c-fos mRNA expression in cortico-striatal-thalamic brain regions. *Neuroscience* **196**, 80–96 (2011).

33. Wilson, R. C. & Niv, Y. Is model fitting necessary for model-based fMRI? *PLoS Comput. Biol.* **11**, e1004237 (2015).

34. Sebold, M. *et al.* Don't think, just feel the music: Individuals with strong Pavlovian-to-instrumental transfer effects rely less on model-based reinforcement learning. *J. Cogn. Neurosci.* **28**, 985–995 (2016).

35. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–47 (1996).

36. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013).

37. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat. Neurosci.* **3**, 1218–1223 (2000).

38. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).

39. Saunders, B. T. & Robinson, T. E. Individual variation in resisting temptation: implications for addiction. *Neurosci. Biobehav. Rev.* **37**, 1955–1975 (2013).

40. Garbusow, M. *et al.* Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict. Biol.* **21**, 719–731 (2016).

41. Schad, D. J. *et al.* Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *Eur. Arch. Psychiatry Clin. Neurosci.* 1–14 (2018). doi:10.1007/s00406-017-0860-4

42. Geurts, D. E., Huys, Q. J. M., den Ouden, H. & Cools, R. Aversive Pavlovian control of instrumental behavior in humans. *J. Cogn. Neurosci.* **25**, 1428–41 (2013).

43. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–6 (1997).

44. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–42 (1997).

45. Garbusow, M. *et al.* Pavlovian-to-instrumental transfer in alcohol dependence: a pilot study. *Neuropsychobiol.* **70**, 111–21 (2014).

46. Wittchen, H.-U. & Pfister, H. *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*. (Swets and Zeitlinger, 1997).

47. R Development Core Team. *R: A language and environment for statistical computing*. (2016).

48. Singmann, H., Bolker, B., Westfall, J. & Aust, F. *afex: Analysis of Factorial Experiments*. (2017).

49. Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. (2018).

50. Ruxton, G. D. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behav. Ecol.* **17**, 688–90 (2006).

51. Canty, A. & Ripley, B. D. *boot: Bootstrap R (S-Plus) functions*. (2017).

52. Davison, A. C. & Hinkley, D. V. *Bootstrap methods and their applications*. (Cambridge University Press, 1997).

53. Morey, R. D. Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason* **4**, 61–4 (2008).

54. Hogarth, L., Dickinson, A. & Duka, T. Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal and uncertainty. *Atten. Assoc. Learn. Brain Behav.* 71–97 (2010).

55. Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R. & Gottlieb, J. Reward modulates attention independently of action value in posterior parietal cortex. *J. Neurosci.* **29**, 11182–91 (2009).

56. Hickey, C., Chelazzi, L. & Theeuwes, J. Reward changes salience in human vision via the anterior cingulate. *J. Neurosci.* **30**, 11096–103 (2010).

57. Hickey, C. & van Zoest, W. Reward creates oculomotor salience. *Curr. Biol.* **22**, R219–20 (2012).

58. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat. Rev Neurosci* **2**, 194 (2001).

59. Gorgolewski, K. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinformatics* **5**, 13 (2011).

60. Iglesias, S. *et al.* Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80**, 519–30 (2013).

61. Deserno, L. *et al.* Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci.* **112**, 1595–600 (2015).

62. White, D. M., Kraguljac, N. V., Reid, M. A. & Lahti, A. C. Contribution of substantia nigra glutamate to prediction error signals in schizophrenia: a combined magnetic resonance spectroscopy/functional imaging study. *Npj Schizophr.* **1**, 14001 (2015).

63. Watanabe, N., Sakagami, M. & Haruno, M. Reward prediction error signal enhanced by striatum-amygdala interaction explains the acceleration of probabilistic reward learning by emotion. *J. Neurosci.* **33**, 4487–93 (2013).

64. Gluth, S., Hotaling, J. M. & Rieskamp, J. The attraction effect modulates reward prediction errors and intertemporal choices. *J. Neurosci.* **37**, 371–82 (2017).

65. Garrison, J., Erdeniz, B. & Done, J. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 1297–310 (2013).

66. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150 (2001).

67. Nebe, S. *et al.* No association of goal-directed and habitual control with alcohol consumption in young adults. *Addict. Biol.* **23**, 379–93 (2018).

68. Neyens, V. *et al.* Representation of semantic similarity in the left intraparietal sulcus: Functional magnetic resonance imaging evidence. *Front. Hum. Neurosci.* **11**, 402 (2017).

69. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–35 (2005).

Acknowledgements

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, FOR 1617: grants SCHA 1971/1-2, HE 2597/13-1, HE 2597/13-2, HE 2597/15-1, SCHL 1969/2-2, SCHL 1969/4-1, SM 80/7-1, SM 80/7-2, WI 709/10-1, WI 709/10-2, ZI 1119/3-1, ZI 1119/3-2, RA 1047/2-1 and RA 1047/2-2; moreover, it was supported in part by DFG CRC-TR 265). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. EF is participant in the BIH Charité Clinician Scientist Program funded by the Charité — Universitätsmedizin and the Berlin Institute of Health. QH acknowledges support by the UCLH NIHR BRC. We thank Nils B. Krömer for helpful feedback and advice on analyses, Sören Kuitunen-Paul for helpful feedback, and Marcus Rothkirch for help with setting up eye-tracking at the Berlin site.

Author contributions

Idea: QJMH; Design: MAR, EF, HUW, USZ, HW, PS, MNS, FS, AH, QJMH; Implementation, pilots, setup: DJS, MG, MS, SN, EO, EF, USZ, MNS, FS, AH and QJMH; Data acquisition: MG, MS, SN, CS with supervision by NRS, HUW, USZ, HW, PS, MNS, FS, AH and QJMH; Analysis: DJS, supervision by MAR, PD, QJMH and input by LD, MRa, FS, and AH; Writing: DJS, MAR, PD, QJMH. All authors read and revised the manuscript and provided critical intellectual contributions.

Competing interests

The authors declare no competing interests.

Figure Legends

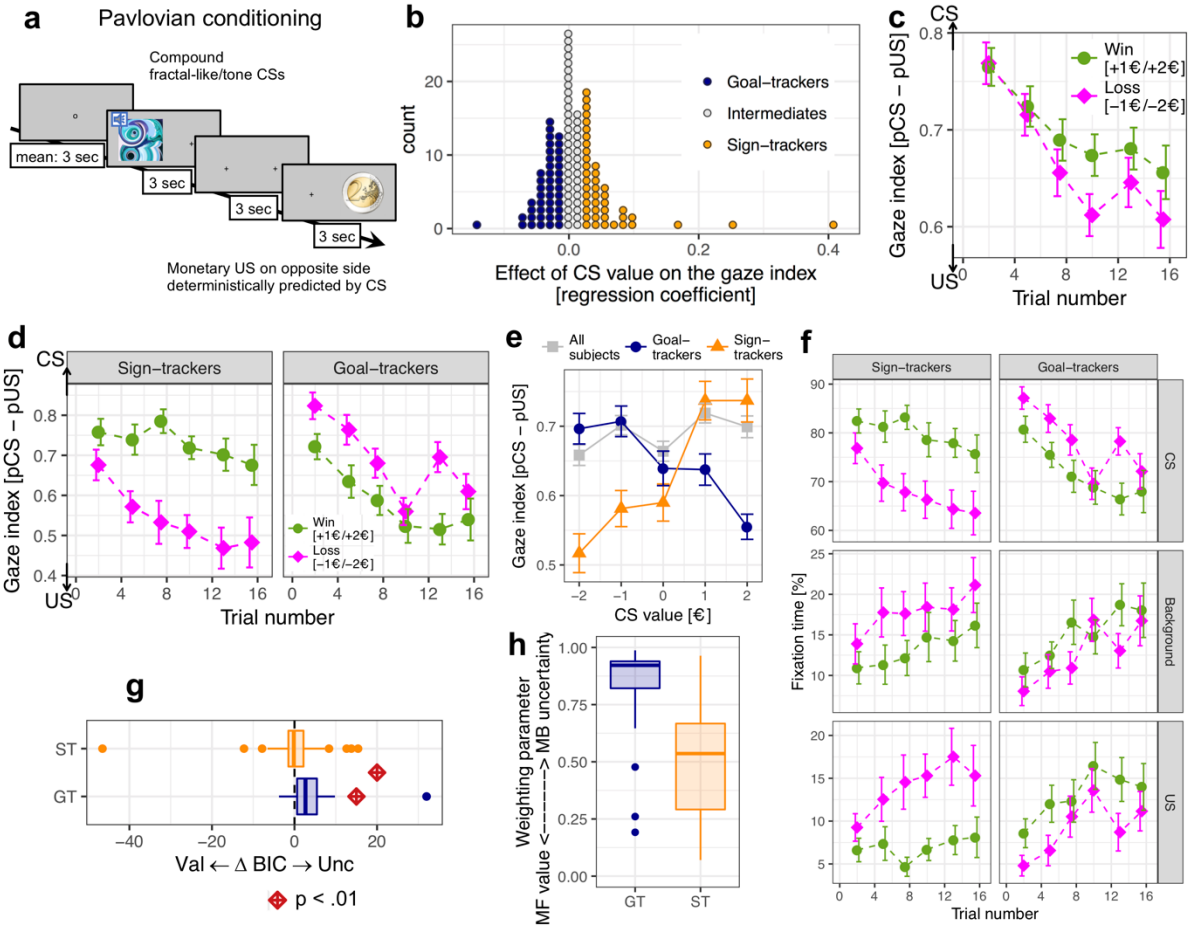

Figure 1. Assessing sign- and goal-trackers via eye-tracking. (a) Pavlovian conditioning paradigm. Conditioned stimuli (CSs) were deterministically followed by positive or negative outcomes (USs). (b-f) Gaze data in the third second of CS presentation. (b) The gaze index captured the tendency to fixate the CS rather than the US. Individual regression coefficients between the gaze index and CS value were broadly distributed around zero. Positive and negative thirds of this distribution were identified as sign- and goal-trackers, respectively. (c) The gaze index was higher for CSs predicting wins than losses ($p_{bootstrap}$ < .05, b = 0.009, $SD_{subjects}$ = 0.057, SE = 0.005, 95% CI = [0.001 0.022], n = 129 subjects), but decreased across trials overall ($p_{bootstrap}$ < .001, b = −0.011, $SD_{subjects}$ = 0.024, SE = 0.002, 99.9% CI = [−0.017 −0.003], n = 129 subjects). The CS value effect on the gaze index increased across trials ($p_{bootstrap}$ < .05, b = 0.0015, $SD_{subjects}$ = 0.0091, SE = 0.0008, 95% CI = [0.00001 0.0032], n = 129 subjects; green vs. magenta lines separate over time). (d) Evolution of gaze index for sign-trackers (left) and goal-trackers (right). (e) Gaze index as a function of CS value. (f) Percentage fixation time on the CS (upper panels), the background (middle panels) and the US location (bottom panels) for CSs predicting wins (green points) and losses (magenta diamonds) across trials in sign-trackers (left panels) and goal-trackers (right panels). (g) Difference in BIC values between computational models of gaze control in sign- and goal-trackers. Positive values indicate support for

model Uncertainty (Unc), which assumes model-based uncertainty controls gaze. Negative values indicate support for model Value (Val), which assumes CS value from a model-free reinforcement learner controls gaze via Pavlovian conditioned responses. For outlier-analyses see Supplementary Information. (h) The computational model parameter ω determines weighting between gaze control by model-free value (ω = 0) versus by model-based uncertainty (ω = 1). Displayed are distributions of the estimated weighting parameter for sign- and goal-trackers. (c-f) Error bars are SEM.

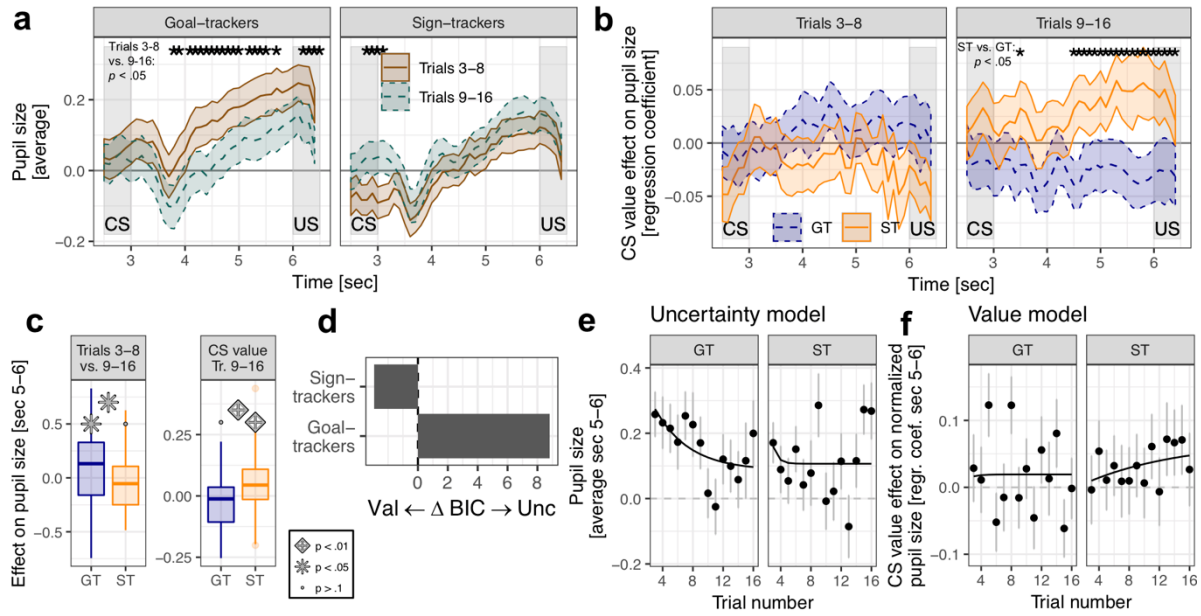

Figure 2. Pupil dilation during Pavlovian conditioning in sign-trackers (ST) and goal-trackers (GT). (a+b) Pupil size between CS and US presentation (seconds 3-6 after CS onset) at the beginning (trials 3-8) and end (trials 9-16) of learning for goal- and sign-trackers. (a) Average pupil size during US anticipation decreases across learning in goal-trackers (t_{140} = −2.29, p = .023, b = −0.055, SE = 0.024, 95% CI = [-0.102 -0.008]) but there was no evidence for a change in sign-trackers (t_{140} = 0.83, p = .405, b = 0.020, SE = 0.025, 95% CI = [-0.028 0.069]; group-difference: F(1, 140) = 4.84, p = .030, η_p^2 = 0.001, 90% CI = [0 0.003]). (b) Sign-trackers show a CS value effect on pupil size (linear regression coefficient) after learning (ST: t_{1314} = 2.89, p = .004, b = 0.521, SE = 0.180, 95% CI = [0.167 0.874], n = 43 subjects), but there is no evidence for the same effect in goal-trackers (GT: t_{1314} = -1.59, p = .112, b = -0.280, SE = 0.176, 95% CI = [-0.625 0.066], n = 43 subjects, ST vs. GT: t_{638} = 3.52, p < .001, b = 0.285, SE = 0.081, 95% CI = [0.126 0.444], n = 86 subjects). (a+b) Stars (*) indicate time points in which the difference is significant (p < .05 from nested tests). (c-f) assess luminance-independent pupil size during seconds 5 to 6. (c) Average pupil size decreases from beginning to end of learning in goal-trackers (left panel; GT: t_{140} = −2.29, p = .023, b = −0.055, SE = 0.024, 95% CI = [-0.102 -0.008], n = 43 subjects), but there is no corresponding evidence in sign-trackers (ST: t_{140} = 0.83, p = .405, b = 0.020, SE = 0.025, 95% CI = [-0.028 0.069], n = 43 subjects; GT vs. ST: F(1, 140) = 4.84, p = .030, η_p^2 =

0.001, 90% CI = [0 0.003], n = 86 subjects). Sign-trackers show a CS value effect on pupil size after learning (right panel; ST: t_{1314} = 2.89, p = .004, b = 0.521, SE = 0.180, 95% CI = [0.167 0.874], n = 43 subjects), but the same effect is not significant in goal-trackers (GT: t_{1314} = -1.59, p = .112, b = -0.280, SE = 0.176, 95% CI = [-0.625 0.066], n = 43 subjects, ST vs. GT: t_{638} = 3.52, p < .001, b = 0.285, SE = 0.081, 95% CI = [0.126 0.444], n = 86 subjects). (d) Difference in BIC values between computational models of pupil dilation in goal- and sign-trackers. Positive values indicate support for model Uncertainty (Unc), which assumes pupil dilation reflects model-based uncertainty. Negative values indicate support for model Value (Val), which assumes pupil dilation reflects learned value from a reinforcement learning model. (e) Average pupil size per trial (points) and pupil size predicted by model Uncertainty (lines) for sign- and goal-trackers. (f) CS value effect on normalized pupil size per trial (points) and CS value effect predicted by model Value (lines) for sign- and goal-trackers. (a-c,e,f) Error bars/bands are SEM.

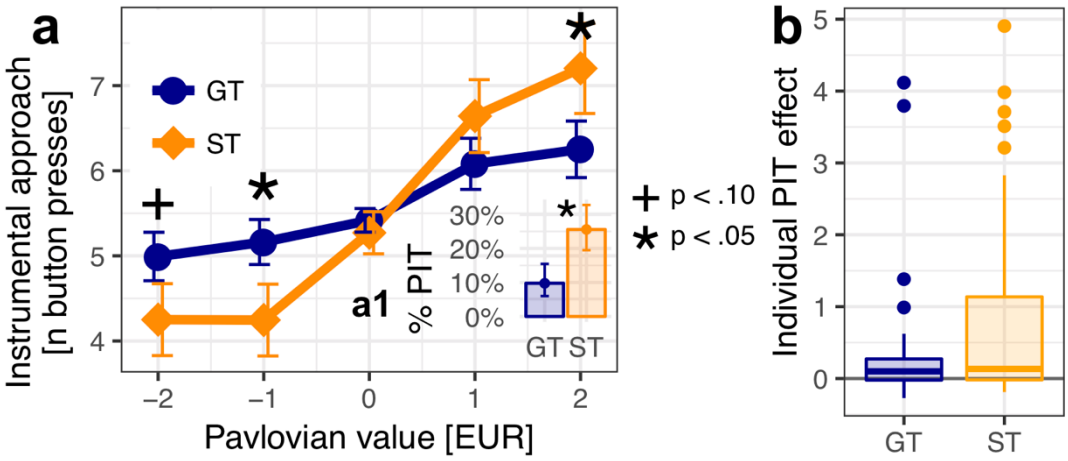

Figure 3. Pavlovian-instrumental transfer (PIT) in sign-trackers (ST) versus goal-trackers (GT). (a) PIT: Instrumental approach increased with the value of Pavlovian CSs in sign-trackers more than in goal-trackers ($p_{bootstrap}$ < .05, b = 0.49, SE = 0.26, 95% CI = [0.09 Inf], n = 84 subjects). (Inset a1) PIT was individually significant in a higher percentage of sign-trackers than goal-trackers ($p_{bootstrap}$ < .05, b = 15.8, SE = 8.3, 95% CI = [1.6 Inf], n = 84 subjects). Error bars are SEM. (b) Distributions of individual PIT effects.

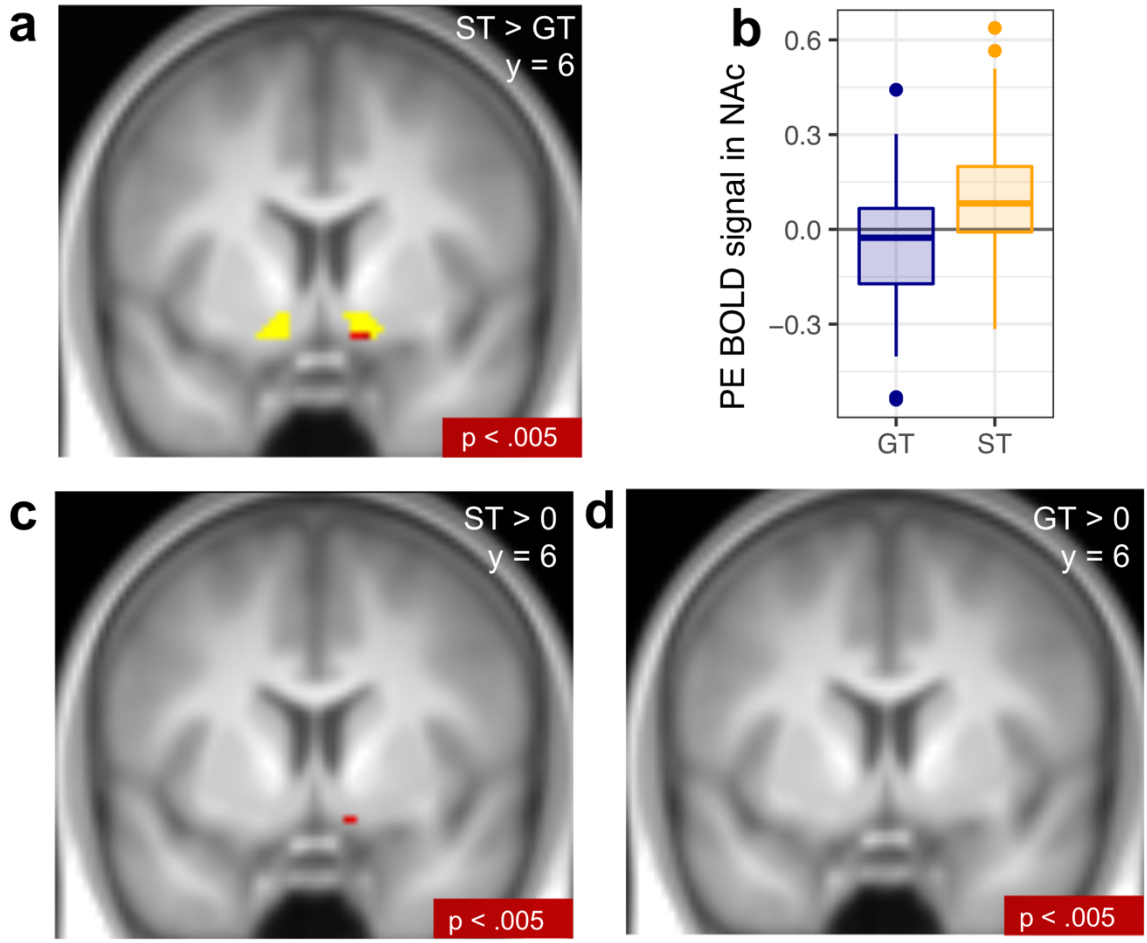

Figure 4. Nucleus accumbens (NAc) BOLD response in sign-trackers (ST) versus goal-trackers (GT). (a) Pavlovian conditioning: A temporal difference reward prediction error (RPE) explains right NAc BOLD response in sign-trackers better than in goal-trackers (red, unmasked; a priori volume of interest, VOI, marked in yellow; $F(1,75) = 10.88$, SVC $p_{FWE} = .026$, [12 6 -14], $\eta_p^2 = 0.122$, 90% CI = [0.031 0.242], n = 78 subjects). (b) RPE signal at the peak response difference in NAc. (c) RPE signal in sign-trackers (red, unmasked; $t_{75} = 3.05$, SVC $p_{FWE} = .025$). (d) RPE signal in goal-trackers. (a+c+d) Threshold: $p < .005$, k = 0.

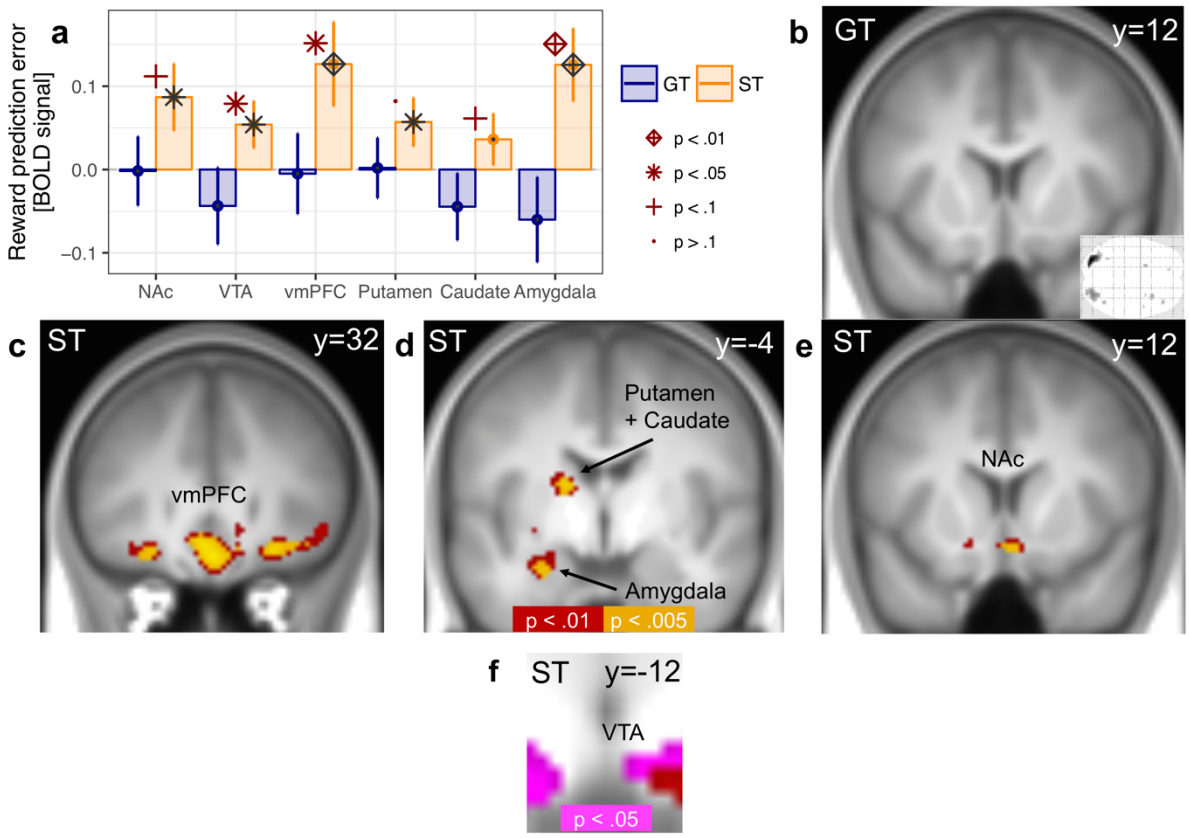

Figure 5. Neural appetitive RPE signals in sign-trackers (ST) versus goal-trackers (GT). (a) The appetitive model-free reward prediction error explains BOLD responses in sign-trackers but not in goal-trackers. Average reward prediction error BOLD response for each volume of interest in sign-trackers (yellow) and goal-trackers (blue). Nucleus accumbens, NAc; ventral tegmental area, VTA; ventromedial prefrontal cortex, vmPFC. Error bars are SEM. (b-f) Voxel-vise results in goal-trackers (b) (see glass brain, inset, $p < .01$, $k = 0$) and in sign-trackers (c-f) in NAc (e), vmPFC (c), Putamen (d), Caudate (d), Amygdala (d), and VTA (f). ST > 0: orange, $p < .005$; red, $p < .01$; pink, $p < .05$; $k = 40$.

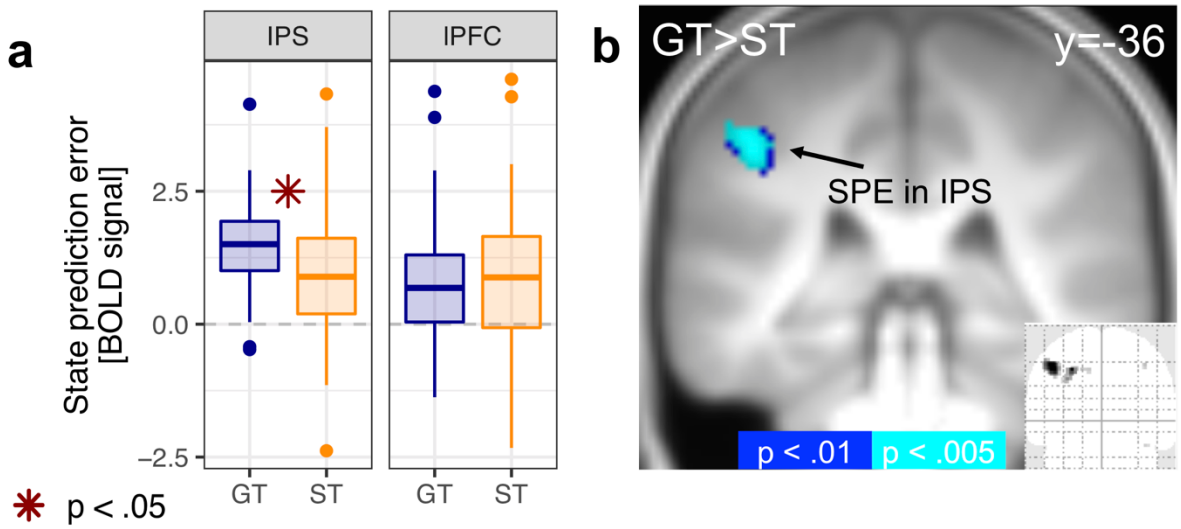

Figure 6. Neural state prediction error learning signals in sign-trackers (ST) versus goal-trackers (GT). (a) Trial-by-trial state prediction error (SPE) predicts BOLD response in intraparietal sulcus (IPS) and lateral prefrontal cortex (IPFC) in goal-trackers (t_{76} = 6.44, p < .001, b = 1.165, SE = 0.181, 95% CI = [0.804 1.53], n = 78 subjects) and sign-trackers (t_{76} = 4.94, p < .001, b = 0.894, SE = 0.181, 95% CI = [0.534 1.25], n = 78 subjects), with a stronger response in goal- than sign-trackers in IPS (t_{67} = 2.12, p = .038, b = 0.564, SE = 0.266, 95% CI = [0.034 1.095], n = 78 subjects). (b) Voxel-vise results for the contrast of a stronger SPE signal in goal- than in sign-trackers (see glass brain, inset, p < .01, k = 0). GT > ST: cyan, p < .005; blue, p < .01; k = 40.