nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0765-5

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[1,2], Elisabeth Obst[3], Christian Sommer[3], Lorenz Deserno[5,6,7], Milena Rabovsky[1], Eva Friedel[2,8], Nina Romanczuk-Seiferth[2], Hans-Ulrich Wittchen[3,9], Ulrich S. Zimmermann[3,10], Henrik Walter[2], Philipp Sterzer[2], Michael N. Smolka[3,11], Florian Schlagenhauf[2,5], Andreas Heinz[2], Peter Dayan[12,13] and Quentin J. M. Huys[14,15,16]

Individuals differ in how they learn from experience. In Pavlovian conditioning models, where cues predict reinforcer delivery at a different goal location, some animals—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 129 male humans using eye-tracking, pupillometry and functional magnetic resonance imaging 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.

Learning from reinforcements involves multiple processes with distinct computational, neural and behavioural signatures. Consider a simple classical Pavlovian conditioning model, where a cue (or conditioned stimulus (CS)) becomes predictive of a reward (or unconditioned stimulus (US)) by being repeatedly presented before the US. In so-called model-free reinforcement learning, learning occurs via reward prediction errors (RPEs[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 that 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 error: so-called state prediction error (SPE[4]), which quantifies 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 this 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, while model-based learning has been suggested to underlie 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 participants 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 CS is presented. Furthermore, sign-trackers, but not goal-trackers will work to obtain the CS after learning[10].

The behavioural differences between sign- and goal-trackers have a number of revealing neural correlates. Sign-trackers learn from RPEs coded in the activity of dopamine neurons[1] and evident in the phasic release of dopamine in the nucleus accumbens (NAc)[11]. These RPE signals initially respond to the rewarding USs, but across learning, shift their response from the US towards the predicting CS. Indeed, sign-trackers depend on the dopaminergic signal to learn, as systemic dopamine blockade disables learning[10].

[1]Cognitive Sciences, University of Potsdam, Potsdam, Germany. [2]Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Charité—Universitätsmedizin Berlin, Berlin, Germany. [3]Department 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]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom. [7]Centre for Computational Psychiatry and Ageing Research, Max Planck University College London, London, UK. [8]Berlin Institute of Health, Berlin, Germany. [9]Institute of Clinical Psyhology & Psychotherapy, TU-Dresden und Department of Psychiatry, Ludwig-Maximalians-Universität, München, Germany. [10]Department of Addiction Medicine and Psychotherapy, kbo-Isar-Amper-Klinikum, Munich, Germany. [11]Neuroimaging Center, Technische Universität Dresden, Dresden, Germany. [12]Gatsby Computational Neuroscience Unit, University College London, London, UK. [13]Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [14]Center for Addictive Disorders, Hospital of Psychiatry, University of Zürich, Zürich, Switzerland. [15]Translational Neuromodeling Unit, University of Zürich and Swiss Federal Institute of Technology, Zürich, Switzerland. [16]Complex Depression Anxiety and Trauma Service, Camden and Islington NHS Foundation Trust, London, UK. *e-mail: danieljschad@gmail.com

That sign-trackers will work to obtain the CS after learning suggests that the dopaminergic RPE underlies a form of learning that 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]. In contrast, for goal-trackers, phasic dopaminergic signals do not evolve with learning, as would be expected from an 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 being predisposed 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 show 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 participants. We hypothesized that the gaze direction during a specifically designed Pavlovian conditioning phase might parallel the behavioural 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, Pavlovian–instrumental transfer (PIT) behaviours and blood-oxygen-level-dependent (BOLD) 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 SPEs, which we investigate here. We found convergent evidence for a double dissociation, with sign-trackers relying on model-free learning and goal-trackers relying on model-based learning.

Results

Participants 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 3 s 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 or €+2), was presented on the other side of the screen. The conditioning task was the second part of a 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 the true CS value for each participant. Figure 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 participants with a positive influence of CS value on the gaze index (yellow in Fig. 1b; $n = 43$). Conversely, we defined goal-trackers as those participants whose gaze approached appetitive US locations[10,15] and avoided aversive US locations (blue in Fig. 1b; $n = 43$). 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 analysing 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 analysis of variance (ANOVA) with the factors location (CS, US and background), CS value (€−2 to €+2) and time from CS onset (1, 2 and 3 s). 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 orientating responses[22,24] insensitive to CS value (no evidence for the interaction CS value × location: $F(8, 2{,}267) = 0.464$; $P = 0.882$; $\eta_p^2 = 0$; 90% confidence interval (CI) = 0 to 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: $t_{2,267} = 3.48$; $P < 0.001$; $b = 0.061$; s.e. = 0.018; 95% CI = 0.027 to 0.096; CS value × location × second 1–3 of CS presentation: $F(11, 1{,}367) = 3.04$; $P < 0.001$; $\eta_p^2 = 0.004$; 90% CI = 0.001 to 0.007) and away from the US location ($t_{2,267} = -1.78$; $P = 0.075$; $b = -0.031$; s.e. = 0.018; 95% CI = −0.066 to 0.003) and background ($t_{2,267} = -1.70$; $P = 0.089$; $b = -0.030$; s.e. = 0.018; 95% CI = −0.065 to 0.005; CS value × location: $F(8, 2{,}267) = 4.458$; $P < 0.001$; $\eta_p^2 = 0.005$; 90% CI = 0.002 to 0.009; Supplementary Fig. 2; for exemplary trials, see Supplementary Fig. 3). This appeared to reflect learning, as this CS value effect increased over trials for the CS location ($t_{2,675} = 2.47$; $P = 0.014$; $b = 0.008$; s.e. = 0.003; 95% CI = 0.002 to 0.015) and decreased for the US location (trend: $t_{2,675} = -1.77$; $P = 0.077$; $b = -0.006$; s.e. = 0.003; 95% CI = −0.013 to 0.0006), but there was no evidence for a change across trials for the background ($t_{2,675} = 0.70$; $P = 0.485$; $b = -0.002$; s.e. = 0.003; 95% CI = −0.009 to 0.004; CS value × location × trial: $F(8, 2{,}675) = 2.94$; $P = 0.003$; $\eta_p^2 = 0.003$; 90% CI = 0.0008 to 0.006). We summarized this effect in the gaze index[10,22] (Fig. 1c), which we analysed 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: $P_{bootstrap} < 0.05$; $b = 0.009$; s.e. = 0.005; 95% CI = 0.001 to 0.022), with the impact of value increasing over trials (interaction of linear fits: CS value × trial number; $P_{bootstrap} < 0.05$; $b = 0.0015$; s.e. = 0.0008; 95% CI = 0.00001 to 0.0032).

Hence, there appeared to be an eye-tracking signal in humans analogous to the behavioural 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: $P_{bootstrap} < 0.001$; $b = 0.059$; s.e. = 0.011; 99.9% CI = 0.039 to 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: $P_{bootstrap} < 0.001$; $b = -0.013$; s.e. = 0.005; 99.9% CI = −0.037 to −0.001; linear CS value × linear trial number: $P_{bootstrap} < 0.05$; $b = 0.003$; s.e. = 0.002; 95% CI = 0.0003 to infinity). 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: $P_{bootstrap} < 0.001$; $b = -0.036$; s.e. = 0.004; 99.9% CI = −0.053 to −0.027), and did so progressively over time.

So far, the definition of goal-tracking is simply the 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

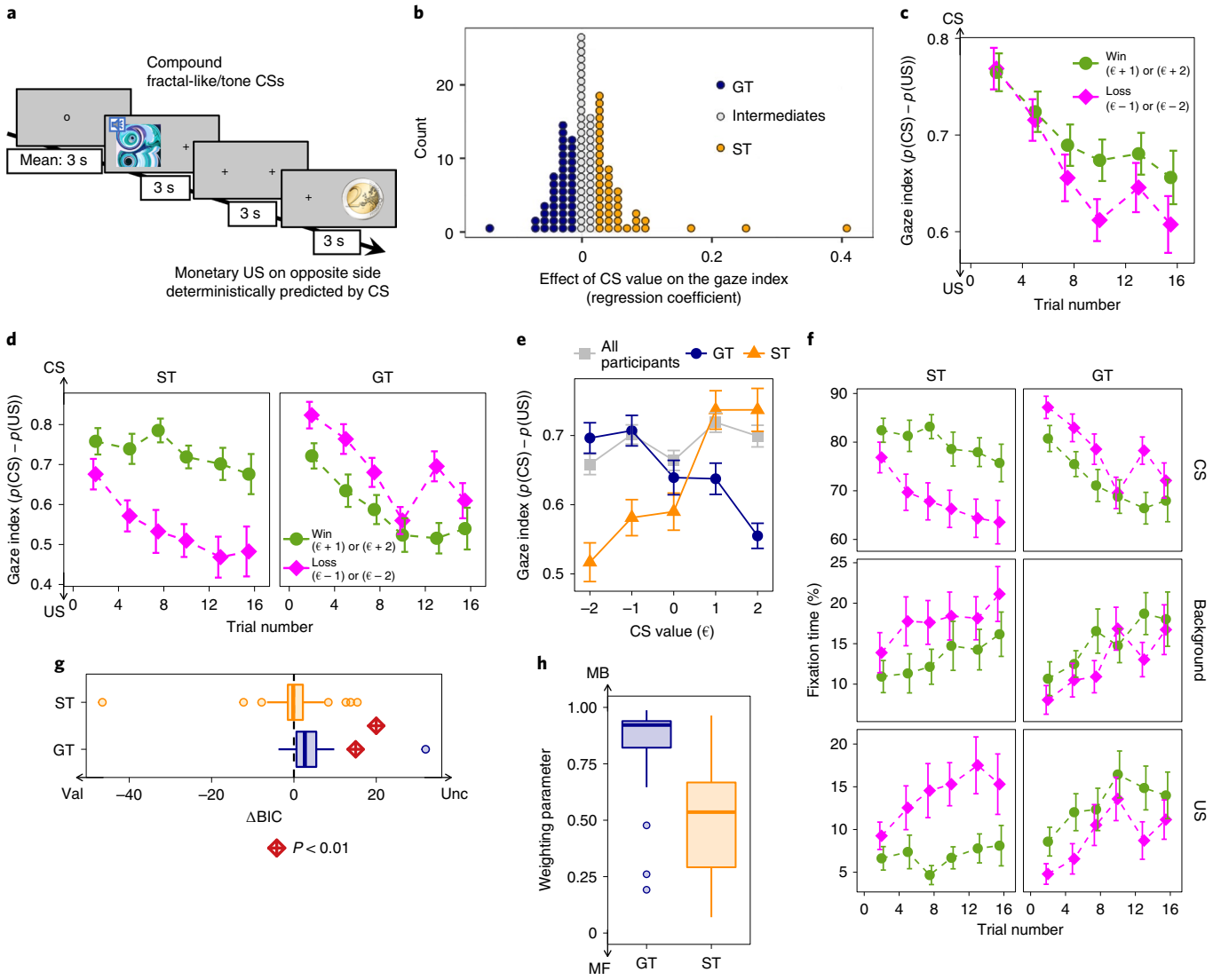

Fig. 1 | Assessment of sign- and goal-trackers via eye-tracking. a, Pavlovian conditioning model. CSs were deterministically followed by positive or negative outcomes (USs). **b**–**f**, Gaze data from the third second of CS presentation. **b**, The gaze index captured the tendency to fixate the CS rather 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-trackers (ST) and goal-trackers (GT), respectively. **c**, The gaze index was higher for CSs predicting wins than for those predicting losses ($P_{bootstrap} < 0.05$; $b = 0.009$; $s.d._{participants} = 0.057$; s.e. = 0.005; 95% CI = 0.001 to 0.022; $n = 129$ participants), but decreased across trials overall ($P_{bootstrap} < 0.001$; $b = -0.011$; $s.d._{participants} = 0.024$; s.e. = 0.002; 99.9% CI = −0.017 to −0.003; $n = 129$ participants). The CS value effect on the gaze index increased across trials, with green versus magenta lines separating over time ($P_{bootstrap} < 0.05$; $b = 0.0015$; $s.d._{participants} = 0.0091$; s.e. = 0.0008; 95% CI = 0.00001 to 0.0032; $n = 129$ participants). **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 (top), background (middle) and the US location (bottom) for CSs predicting wins (green circles) and losses (magenta diamonds) across trials in sign-trackers (left) and goal-trackers (right). **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 that model-based uncertainty controls gaze. Negative values indicate support for model value (val), which assumes that the 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 (MF) value ($\omega = 0$) and model-based (MB) uncertainty ($\omega = 1$). Displayed are the distributions of the estimated weighting parameter for sign- and goal-trackers. In **c**–**f**, error bars are s.e.m. In **g** and **h**, box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits), 1.5× the interquartile range (whiskers) and outliers (points).

independent 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 independent 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 among 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: $t_{410} = -8.62$; $P < 0.001$; $b = -0.008$; s.e. = 0.001; 95% CI = −0.010 to −0.007) towards the US location ($t_{410} = 3.21$; $P = 0.001$; $b = 0.003$; s.e. = 0.001; 95% CI = 0.0012 to 0.0050) and the background ($t_{410} = 5.41$; $P < 0.001$; $b = 0.005$; s.e. = 0.001; 95% CI = 0.003 to 0.007; trial × location: $F(2, 410) = 37.93$; $P < 0.001$; $\eta_p^2 = 0.013$; 90% CI = 0.008 to 0.018). As a result, the gaze index was biased away

from the CS towards the US location (that is, it decreased) with increasing trial number ($P_{bootstrap} < 0.001$; $b = -0.011$; s.e. = 0.002; 99.9% CI = −0.017 to −0.003; Fig. 1c).

To directly test whether this reflected the reduction in uncertainty over the course of training, we implemented computational models assuming that 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 (see Methods). We computed Bayesian information criterion (BIC) values for both models for each participant, and performed a repeated-measures ANOVA with the 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 ($t_{84} = -3.28$; $P = 0.002$; ΔBIC = −3.73; s.e. = 1.14; 95% CI = −6.00 to −1.47; see Fig. 1g), suggesting that state uncertainty drives gaze in goal-trackers. In contrast, the evidence in sign-trackers' gaze was significantly shifted towards the value-based model (model × group: $F(1, 84) = 6.87$; $P = 0.010$; $\eta_p^2 = 0.038$; 90% CI = 0.005 to 0.097), but provided no statistical evidence supporting one model over the other ($t_{84} = 0.43$; $P = 0.672$; ΔBIC = 0.48; s.e. = 1.14; 95% CI = −1.78 to 2.75). Additional modelling, allowing for dual control, where value-based ($\omega = 0$) and uncertainty-based ($\omega = 1$) learning systems within each participant are combined via a weighting parameter (ω), suggested that goal-trackers relied strongly on uncertainty- or model-based control ($\omega_{mean} = 0.84$; $\omega_{s.d.} = 0.18$), whereas sign-trackers seemed to use a mixture of value- and uncertainty-based systems ($\omega_{mean} = 0.48$; $\omega_{s.d.} = 0.24$). We also tested the group difference in the ω parameter using two-tailed non-parametric bootstrapping ($P_{bootstrap} < 0.001$; $b = 0.36$; s.e. = 0.05; 99.9% CI = 0.20 to 0.50; Fig. 1h; see Methods). Hence, examining changes in how individuals freely chose to gaze at a CS or a US allowed us to distinguish two groups of participants who appeared to rely on different computational mechanisms for learning.

The pupil is known to dilate in response to uncertainty, proposedly reflecting noradrenergic arousal signals in the 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 signalled 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 before 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 the factors trial number (3–8 versus 9–16), CS value (€−2 to €+2), time since CS onset (seconds 1–6) and group (sign- versus goal-trackers). Post-hoc tests for second 6 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 (a 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 versus 9–16): $t_{140} = -2.29$; $P = 0.023$; $b = -0.055$; s.e. = 0.024; 95% CI = −0.102 to −0.008). No effect of trial number was observed in sign-trackers ($t_{140} = 0.83$; $P = 0.405$; $b = 0.020$; s.e. = 0.025; 95% CI = −0.028 to 0.069), with a significant group difference ($F(1, 140) = 4.84$; $P = 0.030$; $\eta_p^2 = 0.001$; 90% CI = 0 to 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 with neutral outcomes or losses in the second half of the experiment (trials 9–16), reflecting a value-based pupil response (linear CS value effect: $t_{1,314} = 2.89$; $P = 0.004$; $b = 0.521$; s.e. = 0.180; 95% CI = 0.167 to 0.874). This linear CS value effect developed from the beginning to the end of conditioning (significant increase: $t_{638} = 2.93$; $P = 0.004$; $b = 0.339$; s.e. = 0.116; 95% CI = 0.112 to 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: $t_{1,314} = -1.59$; $P = 0.112$; $b = -0.280$; s.e. = 0.176; 95% CI = −0.625 to 0.066; group difference: $t_{638} = 3.52$; $P < 0.001$; $b = 0.285$; s.e. = 0.081; 95% CI = 0.126 to 0.444). Pupil dilation in goal-trackers therefore appeared to reflect uncertainty[27], while it was driven by CS value in sign-trackers[28]. Next, we 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 participants 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 = 8,023.6; model-free CS value: BIC = 8,032.4; ΔBIC = 8.81; Fig. 2d) and explained the reduction in pupil size across trials seen in goal-trackers only (Fig. 2e). In contrast, in sign-trackers, pupil dilation was best accounted for by model-free CS value (model-based uncertainty: BIC = 7,460.1; model-free CS value: BIC = 7,457.1; ΔBIC = −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.

Next, we attempted to validate the distinction between sign- and goal-trackers in measures independent from eye-tracking. At a behavioural 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 PIT only in sign-trackers[19]. In PIT, appetitive CSs enhance and aversive CSs reduce instrumental approach[21]. The behavioural model employed here contained a PIT phase, in which Pavlovian CSs were presented in the background of the instrumental task; no outcomes were presented, but participants were instructed that outcomes would count towards their reimbursement (Supplementary Fig. 1). We computed the PIT effect for each individual participant 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 goal-trackers. We found that the PIT effect was stronger ($P_{bootstrap} < 0.05$; $b = 0.49$; s.e. = 0.26; 95% CI = 0.09 to infinity; Fig. 3a,b) and more frequently significant at an individual level ($P_{bootstrap} < 0.05$; $b = 15.8$; s.e. = 8.3; 95% CI = 1.6 to infinity; Fig. 3a, inset) in sign-trackers than in goal-trackers, suggesting that the CS acquired incentive salience and elicited Pavlovian approach and avoidance behaviour more in sign-trackers. Second, while sign- and goal-trackers learn differently, they should learn the Pavlovian values equally well. Our model also contained a phase in which participants were forced to choose the better among pairs of CSs and, as expected, sign- and goal-tracker performance was excellent and not statistically different (goal-trackers: 97.8% correct; s.d. = 9.2; sign-trackers: 95.2% correct; s.d. = 14.0; group difference: $P_{bootstrap} > 0.1$; $b = 2.6$; s.e. = 2.6; 95% CI = −1.8 to 8.6; see also Supplementary Fig. 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 RPE response in NAc dopamine concentrations during Pavlovian conditioning[10]. Such RPE signals in human ventral striatum can be measured by 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 the factor group

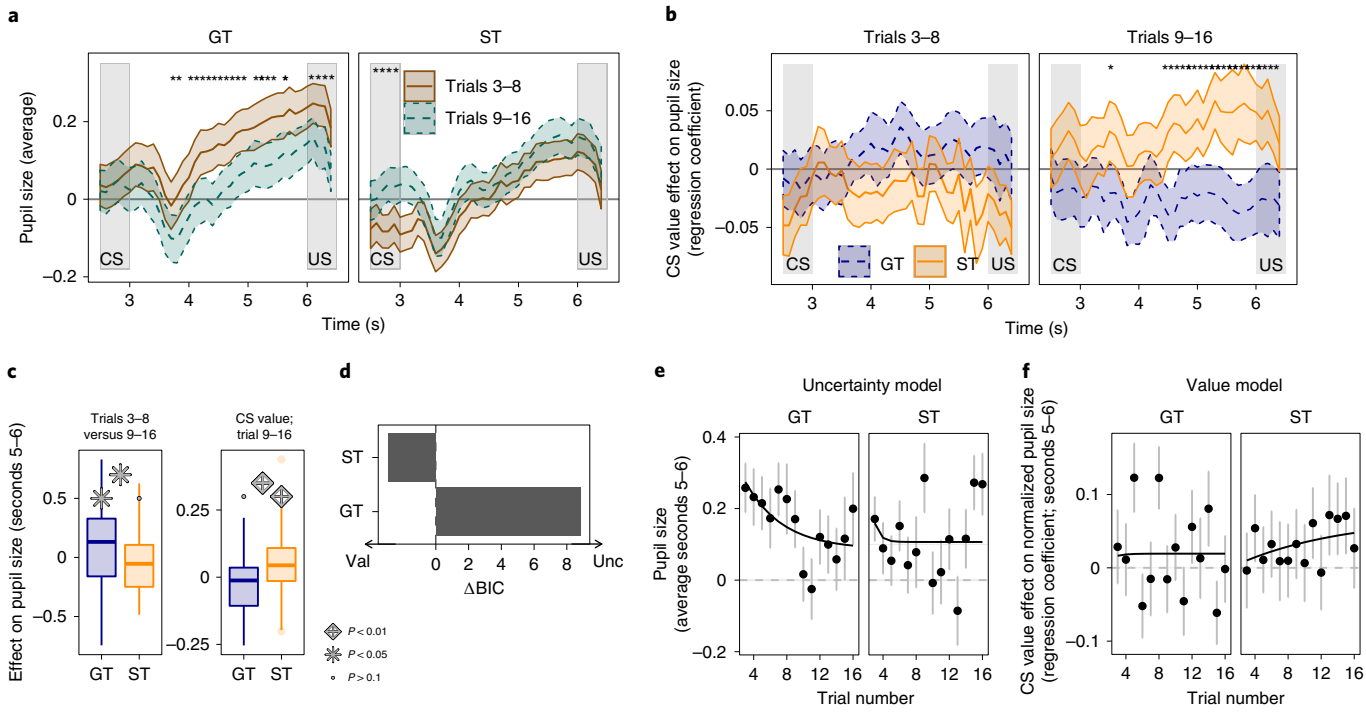

Fig. 2 | Pupil dilation during Pavlovian conditioning in sign-trackers and goal-trackers. a,**b**, Pupil size between CS and US presentation (seconds 3–6 after CS onset) at the beginning and end of learning (stimulus presentations 3–8 and 9–16, respectively) for goal- and sign-trackers. **a**, Average pupil size during US anticipation decreases across learning in goal-trackers ($t_{140}=-2.29$; $P=0.023$; $b=-0.055$; s.e.=0.024; 95% CI=−0.102 to −0.008; $n=43$ participants), but there was no evidence for a change in sign-trackers ($t_{140}=0.83$; $P=0.405$; $b=0.020$; s.e.=0.025; 95% CI=−0.028 to 0.069; $n=43$ participants; group difference: $F(1, 140)=4.84$; $P=0.030$; $\eta_p^2=0.001$; 90% CI=0 to 0.003). **b**, Sign-trackers show a CS value effect on pupil size (linear regression coefficient) after learning ($t_{1,314}=2.89$; $P=0.004$; $b=0.521$; s.e.=0.180; 95% CI=0.167 to 0.874; $n=43$ participants), but there was no evidence for the same effect in goal-trackers ($t_{1,314}=-1.59$; $P=0.112$; $b=-0.280$; s.e.=0.176; 95% CI=−0.625 to 0.066; $n=43$ participants; ST versus GT: $t_{638}=3.52$; $P<0.001$; $b=0.285$; s.e.=0.081; 95% CI=0.126 to 0.444; $n=86$ participants). In **a** and **b**, asterisks indicate time points at which the difference was significant ($P<0.05$ from nested tests). **c**–**f**, Assessment of the luminance-independent pupil size during seconds 5 to 6. **c**, Left: average pupil size decreases from the beginning to the end of learning in goal-trackers ($t_{140}=-2.29$; $P=0.023$; $b=-0.055$; s.e.=0.024; 95% CI=−0.102 to −0.008; $n=43$ participants), but there was no corresponding evidence in sign-trackers ($t_{140}=0.83$; $P=0.405$; $b=0.020$; s.e.=0.025; 95% CI=−0.028 to 0.069; $n=43$ participants; GT versus ST: $F(1, 140)=4.84$; $P=0.030$; $\eta_p^2=0.001$; 90% CI=0 to 0.003; $n=86$ participants). Right: sign-trackers show a CS value effect on pupil size after learning ($t_{1,314}=2.89$; $P=0.004$; $b=0.521$; s.e.=0.180; 95% CI=0.167 to 0.874; $n=43$ participants), but the same effect was not significant in goal-trackers ($t_{1,314}=-1.59$; $P=0.112$; $b=-0.280$; s.e.=0.176; 95% CI=−0.625 to 0.066; $n=43$ participants; ST versus GT: $t_{638}=3.52$; $P<0.001$; $b=0.285$; s.e.=0.081; 95% CI=0.126 to 0.444; $n=86$ participants). Box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits), 1.5× the interquartile range (whiskers) and outliers (points). **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 that pupil dilation reflects model-based uncertainty. Negative values indicate support for model value (val), which assumes that 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. In **a**, **b**, **e** and **f**, error bars/bands represent s.e.m.

(sign- versus goal-trackers) and the covariate testing site to explain the NAc BOLD response. 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: $t_{75}=3.05$; $P_{FWE}=0.025$; Montreal Neurological Institute (MNI) coordinates $[x, y, z]=[12, 6, -14]$), but there was no evidence for such an effect in goal-trackers ($t_{75}=1.58$; SVC $P_{FWE}=0.398$), with a significant group difference ($F(1, 75)=10.88$; SVC $P_{FWE}=0.026$; MNI coordinates [12, 6, −14], $\eta_p^2=0.122$; 90% CI=0.031 to 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 analysis testing the RPE for wins (€0, €+1 and €+2) and losses (€0, €−1 and €−2) separately. For this analysis, we extracted the average appetitive or aversive RPE BOLD signal within the a priori NAc VOI for each participant, and performed one-sample *t*-tests of the hypotheses that the appetitive RPE signal in sign-trackers is larger than zero, and that it is 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 ($t_{38}=2.15$; $P=0.019$; $b=0.087$; s.e.=0.040; 95% CI=0.019 to infinity; Fig. 5a,e), but no NAc RPE response in goal-trackers (Fig. 5a,b; $t_{38}=-0.04$; $P=0.516$; $b=-0.002$; s.e.=0.042; 95% CI=−0.072 to infinity; group difference: $t_{76}=1.53$; $P=0.065$; $b=0.089$; s.e.=0.058; 95% CI=−0.008 to infinity; Supplementary Fig. 5). However, the aversive RPE involving losses did not elicit BOLD responses ($P>0.1$; see Supplementary Information).

RPE(-like) signals are also found in other brain regions, such as ventral tegmental area (VTA)/substantia nigra, dorsal striatum (caudate and putamen), ventromedial prefrontal cortex (vmPFC)

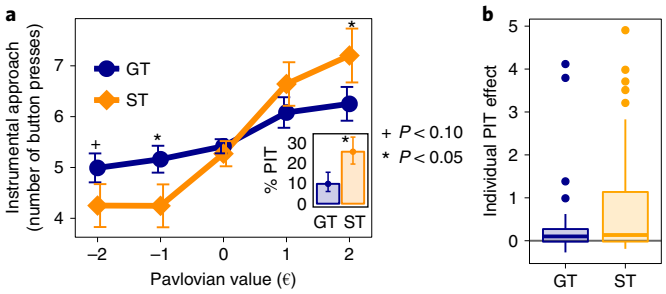

Fig. 3 | PIT in sign-trackers versus goal-trackers. a, Instrumental approach increased with the value of Pavlovian CSs in sign-trackers more than in goal-trackers ($P_{bootstrap} < 0.05$; $b = 0.49$; s.e. = 0.26; 95% CI = 0.09 to infinity; $n = 84$ participants). Inset: PIT was individually significant in a higher percentage of sign-trackers than goal-trackers ($P_{bootstrap} < 0.05$; $b = 15.8$; s.e. = 8.3; 95% CI = 1.6 to infinity; $n = 84$ participants). Error bars represent s.e.m. **b**, Distributions of individual PIT effects. Box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits), 1.5× the interquartile range (whiskers) and outliers (points).

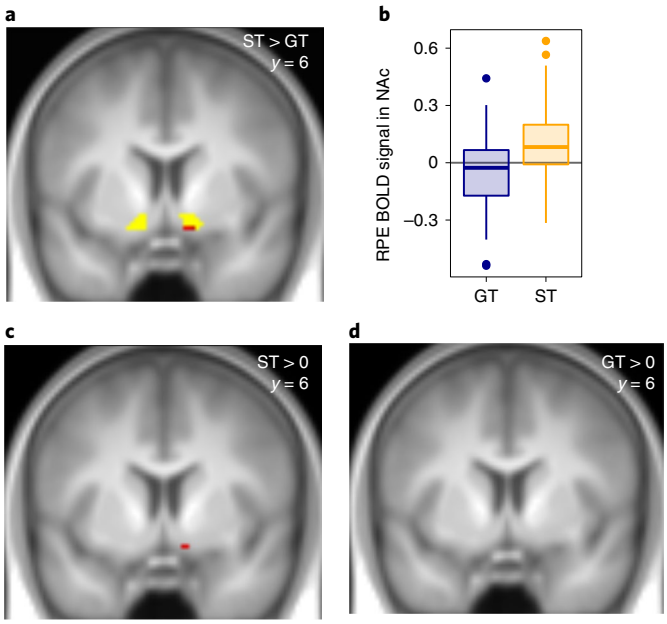

Fig. 4 | NAc BOLD response in sign-trackers versus goal-trackers. **a**, Pavlovian conditioning: a temporal-difference reward prediction error explains the right NAc BOLD response in sign-trackers better than in goal-trackers (red, unmasked; a priori VOI marked in yellow; $F(1, 75) = 10.88$; SVC $P_{FWE} = 0.026$; MNI coordinates [12, 6, -14]; $\eta_p^2 = 0.122$; 90% CI = 0.031 to 0.242; $n = 78$ participants). $y = 6$ indicates the illustrated plane in MNI coordinates. **b**, RPE signal at the peak response difference in NAc. Box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits), 1.5× the interquartile range (whiskers) and outliers (points). **c**, RPE signal in sign-trackers (red, unmasked; $t_{75} = 3.05$; SVC $P_{FWE} = 0.025$). **d**, RPE signal in goal-trackers. In **a**, **c** and **d**, the threshold P value was < 0.005 (cluster size $k = 0$).

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 ref. [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 the factors VOI (VTA/substantia nigra, caudate, putamen, vmPFC and 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 ($t_{76} = 1.89$; $P = 0.031$; $b = 0.035$; s.e. = 0.019; 95% CI = 0.004 to infinity), and these were stronger than in goal-trackers throughout ($F(1, 76) = 4.18$; $P = 0.044$; $\eta_p^2 = 0.01$; 90% CI = 0 to 0.03; RPE signal in goal-trackers: $t_{76} = -1.00$; $P = 0.322$; $b = -0.019$; s.e. = 0.019; 95% CI = −0.049 to infinity). There was no evidence that the group difference differed between VOIs ($F(3, 241) = 1.07$; $P = 0.363$; $\eta_p^2 = 0.003$; 90% CI = 0 to 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 an RPE signal ($t_{76} = 2.38$; $P = 0.010$; $b = 0.080$; s.e. = 0.034; 95% CI = 0.024 to infinity), which was stronger than in goal-trackers ($F(1, 76) = 5.40$; $P = 0.023$; $\eta_p^2 = 0.014$; 90% CI = 0.001 to 0.039), where goal-trackers showed no evidence for an RPE signal ($t_{76} = -0.90$; $P = 0.369$; $b = -0.030$; s.e. = 0.034; 95% CI = −0.086 to infinity). We explored appetitive RPE signals in individual VOIs (Fig. 5a,c–f and Supplementary Fig. 6). We found that appetitive RPE signals in sign-trackers were stronger than in goal-trackers in several VOIs (VTA: $P = 0.038$; vmPFC: $P = 0.032$; putamen: $P = 0.121$; caudate: $P = 0.058$; amygdala: $P = 0.004$), and that only the effect in the amygdala ($P = 0.020$), but not in the other VOIs ($P > 0.1$), survived correction for the multiple exploratory tests. However, these differences should be interpreted with caution given that 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 behaviour[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 ($P = 0.034$), 0.2 ($P = 0.039$), 0.3 ($P = 0.046$), 0.4 ($P = 0.048$), 0.5 ($P = 0.0495$) and 0.6 ($P = 0.04998$). 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 few data exist[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 intraparietal sulcus (IPS) and in the lateral PFC (lPFC)[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 the factors VOI (IPS 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 IPS. We found SPE signals in both the IPS and the lPFC (Fig. 6a), which were significant in both goal-trackers ($t_{76} = 6.44$; $P < 0.001$; $b = 1.165$; s.e. = 0.181; 95% CI = 0.804 to 1.53) and sign-trackers ($t_{76} = 4.94$; $P < 0.001$; $b = 0.894$; s.e. = 0.181; 95% CI = 0.534 to 1.25; also see Supplementary Fig. 8), consistent with the behavioural signatures showing a model-based component in both groups. However, in the IPS, this SPE signal was stronger in goal- than in sign-trackers ($t_{67} = 2.12$; $P = 0.038$; $b = 0.564$; s.e. = 0.266; 95% CI = 0.034 to 1.095; interaction group × VOI: $F(1, 76) = 5.34$; $P = 0.024$; $\eta_p^2 = 0.033$; 90% CI = 0.002 to 0.092; Fig. 6a,b and Supplementary Fig. 9). As previous work[4] has also shown this area to relate to behaviour, this difference may underlie goal-trackers' stronger reliance on model-based Pavlovian learning in the current task.

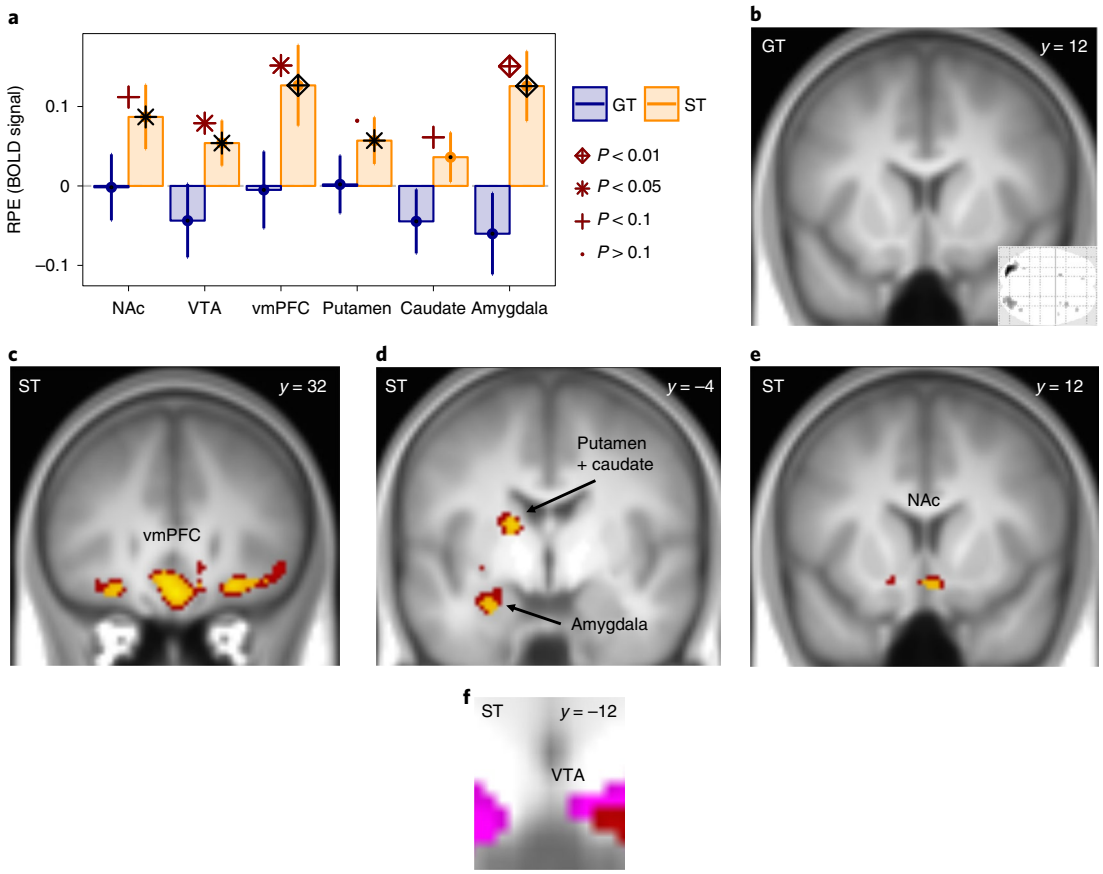

Fig. 5 | Neural appetitive RPE signals in sign-trackers versus goal-trackers. a, The appetitive model-free RPE explains BOLD responses in sign-trackers but not in goal-trackers. Average RPE BOLD responses for each VOI in sign-trackers (yellow) and goal-trackers (blue) are shown. Error bars represent s.e.m. **b**–**f**, Voxel-vise results in goal-trackers (**b**) (inset: glass brain; $P < 0.01$; $k = 0$) and in sign-trackers (**c**–**f**) in the vmPFC (**c**), putamen (**d**), caudate (**d**), amygdala (**d**), NAc (**e**) and VTA (**f**). For RPE > 0 in ST, $P < 0.005$ (orange), $P < 0.01$ (red) or $P < 0.05$ (pink) ($k = 40$).

Finally, if it is true that the imaging and eye measures relate to the same learning processes, 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 $\eta_{\text{pupil}} = 0.19$, which was well in line with the learning rate of $\eta_{\text{fMRI}} = 0.15$ used for fMRI analyses. The value signal in sign-trackers yielded a learning rate for value of $\alpha_{\text{pupil}} = 0.06$, which was comparable with the learning rate showing the strongest NAc RPE signal in model-based fMRI analyses ($\alpha_{\text{fMRI}} = 0.05$; 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. Here, we 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 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 contrast, in goal-trackers, gaze and pupil size were related to model-based uncertainty. This was accompanied by a stronger model-based SPE signal in the IPS. Goal-trackers also showed reduced 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. Hence, group differences, as in animals[10], may not reflect differences in learning ability, but rather indicate different mechanisms or systems underlying learning. Taken together, eye-tracking, pupillometry, behavioural 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 and caudate), VTA, amygdala and vmPFC, in the sign-trackers, while we found no evidence for RPE signals in these areas in the goal-trackers. Early theories (for example, ref. [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

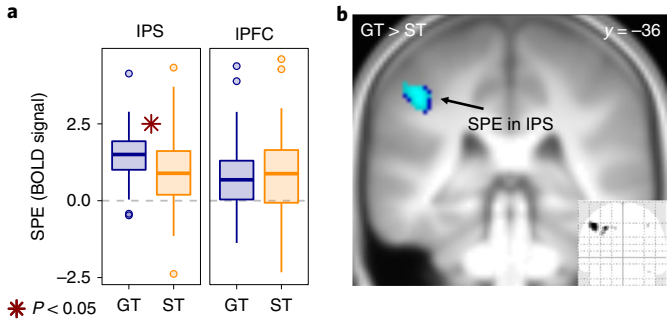

Fig. 6 | Neural SPE learning signals in sign-trackers versus goal-trackers. **a**, Trial-by-trial SPE predicts BOLD response in the IPS and lPFC in goal-trackers ($t_{76}=6.44$; $P<0.001$; $b=1.165$; s.e.$=0.181$; 95% CI$=0.804$ to 1.53; $n=39$ participants) and sign-trackers ($t_{76}=4.94$; $P<0.001$; $b=0.894$; s.e.$=0.181$; 95% CI$=0.534$ to 1.25; $n=39$ participants), with a stronger response in goal- than sign-trackers in the IPS ($t_{67}=2.12$; $P=0.038$; $b=0.564$; s.e.$=0.266$; 95% CI$=0.034$ to 1.095; $n=78$ participants). Box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits), 1.5× the interquartile range (whiskers) and outliers (points). **b**, Voxel-vise results for the contrast of a stronger SPE signal in goal- than in sign-trackers. Inset: glass brain ($P<0.01$; $k=0$). For GT > ST, $P<0.005$ (cyan) or $P<0.01$ (blue) ($k=40$).

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 and attention, and through pupil dilation, also arousal. While outcome-specific PIT may depend in part on model-based mechanisms, the current model 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 behaviour.

While all of 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 IPS[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 (that is, associability)[37] and pupil dilation[27], and is thought to be more resistant to Pavlovian response biases compared with 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 show a double dissociation between model-free RPE learning mediating incentive salience attribution in sign-trackers and 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 a 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 were probably 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]. However, computational modelling 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, behaviour 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 model) when removing half of 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 participants, 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: (1) we analysed brain activity in regions known to encode prediction error signals; and (2) we studied 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 that they were not driven by differences in visual input, but were due to differences in prediction error signalling.

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

This research complied with all of the 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), and procedures were in accordance with the Declaration of Helsinki. Informed consent was obtained from all human participants. Participants received a monetary compensation of €10 h^{-1} 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 of fixation time on the CS minus the proportion of fixation time on the US location during the third second of CS presentation (that is, gaze index = p(CS) – p(US)). A gaze index of 1 indicated that 100% of fixation time was spent on the CS; a gaze index of −1 indicated that 100% of fixation time was spent on the US; a gaze index of 0 indicated the same percentage of fixation time on the CS and US; and intermediate values indicated intermediate fixation statistics. For each individual participant, we computed a linear regression of gaze index on the true value of the CS (that is, €−2, €−1, €0, €+1 or €+2). Participants for whom gaze was attracted more to win-predictive than to loss-predictive CSs had a positive regression coefficient of CS value, and we defined the third of participants with the most positive regression coefficients as sign-trackers (n = 43). Participants for whom gaze was attracted towards the goal for expected wins more than for expected losses had a negative regression coefficient, and we defined goal-trackers as the third of participants with the most negative regression coefficient (n = 43).

Overview of computational modelling. Model-based valuation estimates the probabilities of 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: $T_{i,j} = p(\mathrm{US}_j|\mathrm{CS}_i)$. Moreover, it estimates a 'reward matrix', which stores expected reward R for each US outcome j: $R_j = E[R|\mathrm{US}_j]$. When a CS i is presented, the model-based system determines the expected value $\mathbf{V}_i^{\mathrm{MB}}$ by considering all of the possible US outcomes, and by weighing their expected rewards by their transition probabilities. This can be formulated by multiplying the transition matrix with the reward matrix: $V_i^{\mathrm{MB}} = T_i \times \mathbf{R}$. 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 (that is, $T_{i,j}^0 = \frac{1}{5} = 0.2$ *for all* i, j). Participants experience an SPE when they encounter a transition from a CS i to a US j in trial t, $\delta_t^{\mathrm{SPE}} = 1 - T_{i,j}^t$. They use this to update the transition matrix, $T_{i,j}^{t+1} = T_{i,j}^t + \eta \times \delta_t^{\mathrm{SPE}}$. To keep the matrix normalized to total probabilities of one, transition probabilities for the other US $j\prime \neq j$ that have not been observed are reduced by $T_{i,j\prime}^{t+1} = T_{i,j\prime}^t \times (1 - \eta)$. We assume that the reward matrix is known instantly and with certainty. We approximate uncertainty (unc) in state predictions as $\mathbf{unc}_t = 1 - \max_j T_j^t$. We repeated core analyses with a fully Bayesian model-based learner with explicit uncertainty computation, with comparable results.

In contrast, model-free learning 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 RPEs, $\delta_t^{\mathrm{RPE}} = \mathbf{R}^t - \mathbf{V}_{i,t}^{\mathrm{MF}}$, where R^t is the experienced reward value in trial t and $V_{i,t}^{\mathrm{MF}}$ is the current model-free state value for CS i in trial t, to update estimates of expected value directly: $V_{i,t+1}^{\mathrm{MF}} = V_{i,t}^{\mathrm{MF}} + \alpha \times \delta_t^{\mathrm{RPE}}$.

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 $\widehat{\mathbf{y}_t^{\mathrm{ID}}} = \mathrm{int}_{\mathrm{unc}} + b_{\mathrm{unc}} \times \mathbf{unc}_t$, where $\widehat{y_t^{\mathrm{ID}}}$ of participant 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, $\mathrm{int}_{\mathrm{unc}}$ is a free intercept or baseline parameter capturing the expected value of the dependent variable after learning and complete state certainty, and b_{unc} is a free parameter for the weight of model-based uncertainty, which we reparameterize to $b_{\mathrm{unc}} = e^{b_{\mathrm{unc}}\prime}$, and which measures the degree to which maximum 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 $\widehat{\mathbf{y}_t^{\mathrm{ID}}} = \mathrm{int}_{\mathrm{val}} + b_{\mathrm{val}} \times \mathbf{V}_t^{\mathrm{MF}}$, where $\mathrm{int}_{\mathrm{val}}$ is a free intercept or baseline parameter capturing the average gaze direction or pupil size, and b_{val} is a free parameter measuring the weight of model-free value influencing gaze or pupil size (that is, the Pavlovian response bias).

For gaze direction and pupil size, we assume the likelihood for individual observations y_t is normally distributed $P(\mathbf{pupil}|\hat{\mathbf{y}}, \sigma^2) = \prod_{\mathrm{ID}} \prod_t P\left(\mathbf{y}_t^{\mathrm{ID}}|\widehat{\mathbf{y}_t^{\mathrm{ID}}}, \sigma^2\right)$, where y_t is the observed dependent variable per trial, $\hat{y}_t$ 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 $\mathbf{gaze}^{\mathrm{ID}} = \left\{\mathbf{y}_t^{\mathrm{ID}}\right\}_{t=1}^{T}$ per participant ID across T trials is then $P\left(\mathbf{gaze}^{\mathrm{ID}}|\widehat{\mathbf{y}^{\mathrm{ID}}}, \sigma_{\mathrm{ID}}^2\right) = \prod_t P\left(\mathbf{y}_t^{\mathrm{ID}}|\widehat{\mathbf{y}_t^{\mathrm{ID}}}, \sigma_{\mathrm{ID}}^2\right)$. For the pupil size, $\mathrm{pupil} = \left\{\left\{\mathbf{y}_t^{\mathrm{ID}}\right\}_{t=1}^{T}\right\}_{\mathrm{ID}=1}^{n}$ of n participants with each T trials, we pool the likelihood across all participants:

$$P(\mathbf{pupil}|\hat{\mathbf{y}}, \sigma^2) = \prod_{\mathrm{ID}} \prod_t P\left(\mathbf{y}_t^{\mathrm{ID}}|\widehat{\mathbf{y}_t^{\mathrm{ID}}}, \sigma^2\right)$$

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 the model-free value expectation from 0 (at trial onset) to the predictive value of the CS, $V_{i,t}^{\mathrm{MF}}$, yielding a temporal-difference RPE of $\delta_{\mathrm{CS}}^{\mathrm{RPE}} = V_{i,t}^{\mathrm{MF}}$. At US onset, value expectation changes from the predictive value of the CS, $V_{i,t}^{\mathrm{MF}}$, to the observed US value, R^t (that is, $\delta_{\mathrm{US}}^{\mathrm{RPE}} = \mathbf{R}^t - V_{i,t}^{\mathrm{MF}}$)[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 $\delta_t^{\mathrm{SPE}} = 1 - T_{i,j}^t$ to modulate fMRI activity at the time of US onset. The RPE and 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 s of stimulus presentation.

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

Task. The task tested 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 the scanning session and the forced-choice task was conducted after it. Pavlovian conditioning and PIT were assessed during fMRI. The task was programmed using MATLAB 2011 (MATLAB version 7.12.0; MathWorks) with the Psychophysics Toolbox Version 3 extension[43,44]. It was presented on a computer screen (instrumental training and forced choice) and on a projector via a mirror system (Pavlovian conditioning and PIT). For a detailed description of the task, see refs. [40,45].

Instrumental conditioning. Participants were instructed to collect or avoid shells by repeated button presses. To collect a shell, participants 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 (2-s response window) were needed to collect a shell, which participants were not informed about. Participants received probabilistic feedback. On approach trials, a 'good' shell was monetarily rewarded in 80% of trials and punished in 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% of trials and rewarded in 20% of trials, and vice versa if not collected. Participants learned to respond to three 'good' (that is, approach) and three 'bad' (that is, non-approach) shells through trial and error. Participants performed 60–120 instrumental training trials, depending on their performance. To ensure that all of the participants 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 3 s. This was followed by a delay of 3 s with two fixation crosses at the two potential CS locations (left and right; Supplementary Fig. 1b). Finally, the US was presented for 3 s at the position opposite 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 participants. 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.

Participants were instructed to observe the CSs and USs and to memorize the pairings. All participants 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. Participants then performed 90 trials of the instrumental task, as in training, but with fractal CSs tiling the background (Supplementary Fig. 1c). No outcomes were presented, but participants were instructed that their choices still counted towards 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 × 5 × 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 (see refs. [17,21]).

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

Forced-choice task. Finally, participants 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 ten 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 2 s 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 of whom were male and 18 years old.

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 (Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)[46] axis I and Composite International Diagnostic Interview (CIDI)[47]) and neurological disorders.

Here, we report on a subsample of 144 participants for whom eye-tracking data were available during Pavlovian conditioning. These same participants were tested on all tasks (that is, we tested the same sample repeatedly). For each task and recording technique, data for some participants were missing for technical reasons (for example, recording failures or early task abortion). For numbers of missing data, see below.

The study was designed, and a sample size of 198 participants 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 3 years. For our cross-sectional analysis, valid eye-tracking data in the fMRI scanner were available for 129 participants. The statistical group identification (described in the manuscript) resulted in a final sample size of 43 participants per sign-/goal-tracker group, which yielded 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 the statistical tests described in the Methods. Assignment of experimental stimuli (Pavlovian CSs and 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 the 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 1,000 Hz; in Dresden) and 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, participants were instructed to fixate on a central fixation point. Failure to fixate led 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: 2,410 ms; echo time: 25 ms; flip angle: 80°; field of view: $192\times192\,mm^2$; voxel size: $3\times3\times2\,mm^3$) comprising 42 slices at 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: 1,900 ms; echo time: 5.25 ms; flip angle: 9°; field of view: $256\times256\,mm^2$; 192 sagittal slices; voxel size: $1\times1\times1\,mm^3$). Before functional scanning, a field map was collected to account for individual homogeneity differences of the magnetic field.

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

Data analyses and statistics. Data were analysed using MATLAB 2013a (MATLAB version 8.1.0.604; MathWorks) and the R System for Statistical Computing[48]. fMRI data were analysed using Statistical Parametric Mapping 8 (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[49]. Contrasts were computed using the R package emmeans[50]. For fMRI analyses, at the second level, we performed either one-sample Student's *t*-tests or two-sample Welch's *t*-tests (capturing situations of equal and unequal variances)[51]. For random effects analyses of behavioural 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 confidence intervals (90, 95, 99 and 99.9%; R package boot[52,53]), was used instead. Statistical tests were two-tailed unless otherwise indicated. Error bars in the figures represent repeated-measures s.e.m.[54] unless otherwise indicated. Error bars for pupil analyses (Fig. 2e,f) were extracted via linear mixed-effects models. Box-and-whisker plots show the median (centre line), upper and lower quartiles (box limits); 1.5× the interquartile range (whiskers) and outliers (points). For *F*-tests, as a measure of the effect size, we report the proportion of variance of the dependent variable accounted for by the levels of the factor (that is, n^2_p), together with 95% CIs, as computed by the function ci.pvaf() from the R package MBESS[55].

Eye-tracking analyses of gaze. *Preprocessing.* Data from the EyeLink 1000 system (right eye) were downsampled to the 50 Hz that was available for the iViewX system. Given different sampling rates between testing sites, we checked for differences in the gaze-index results (Fig. 1b–d and Supplementary Fig. 2a), but found no significant difference ($P>0.1$). We corrected for temporal–spatial drifts and distortions of the eye-tracking data for each participant and each calibration across trials. A total of 15 participants showed poor correction performance and were removed from the analysis after visual inspection, yielding 129 participants with valid eye-tracking data. We repeated some of the core analyses using uncorrected eye-tracking data, and found overall consistent results.

No valid gaze data were recorded after the second calibration in three out of the 129 valid participants (one sign-tracker and two controls). In one (sign-tracker) participant, no eye-tracking data were available for the last 22 trials. Additionally, an average of 12.8% of the eye-tracking samples recorded during CS presentation (median = 10.6%; s.d. across participants = 10.0%) were 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%; s.d. = 2.2%) trials with no valid gaze data overall, and for the third second of CS presentation, an average (s.d.) of 3.5 (6.6)% of trials with no valid gaze data.

Valid gaze samples were classified as being directed at one of three spatial regions of interest (ROIs): (1) the CS; (2) the spatial location of later US presentation; and (3) the rest of the screen reflecting the background. Note that for each second of CS presentation, the percentage of samples within each ROI (p(ROI)) also reflects the cumulative gaze times (that is, dwell times) for these ROIs, proportionally corrected for missing/invalid data.

Percentage fixation times. CS onset is known to trigger initial orientating responses[22,56] that do not differ between sign-trackers and goal-trackers[24]. Later on, gaze exhibits Pavlovian conditioned responses to CS value, visible in enhanced fixations on appetitive compared with neutral or aversive cues[22,56–58] (for early responses, see ref. [59]). 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 or €−2) on percentage fixation times on the CS, US location and background differed between the 3 s 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 centred). Regression coefficients were analysed via repeated-measures ANOVA to test how trial effects differed between the CS, US location and background (factor location), between CS value levels (€+2, €+1, €0, €−1 and €−2) and between the 3 s 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 × trial number × location for the third second of CS presentation. We used planned contrasts to test linear CS value effects (€−2, €−1, €0, €+1 or €+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 the relative approach to a CS and US[10]. The gaze index was 1 if the entire time was spent looking at the CS, −1 if the entire time was spent looking at the US, and 0 if there was no preference for either the CS or the US location, and it had some intermediate value if gaze was distributed between the CS, US location and 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 became stronger across trials (interaction CS value × trial number per stimulus; refs. [19,22,37,57]).

Definition of sign-tracker and goal-tracker based on gaze index. To define sign-trackers versus goal-trackers, we computed the influence of CS value on the gaze index during the third (that is, last) second of CS presentation per participant. Sign-trackers were defined as the third with the most positive regression coefficients ($n=43$), while goal-trackers were defined as the third with the most negative ($n=43$). We tested whether the frequency of sign-trackers versus goal-trackers differed between testing sites (Berlin/Dresden) using a chi-squared test.

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

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 $T_0(\mathrm{CS,US}) = 1/5$ as all USs are equally likely. In each conditioning trial t, the model-based system computes an SPE:

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

and updates the probability $T(\mathrm{CS,US})$ of the observed transition via:

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

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

We approximate uncertainty (U) in state predictions in this experience- and model-based system via the distance of the highest state transition probability conditional on the visited CS 'cs' from certainty (that is $U = 1 - \max[T(\mathrm{CS = cs,US})]$). We assumed that predictive uncertainty directly increases attention towards the predictive CS, and hence increases CS-related eye fixations. We therefore modelled the trial-by-trial gaze index via:

$$\text{Model MB}: \mathbf{GazeIndex}_t = c + \beta_U^{\mathrm{gaze}} \times \mathbf{U}_t(s) \quad (3)$$

where 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 (for example, capturing effects of visual salience[60]), $U_t(s)$ is the trial-by-trial CS-related uncertainty, and β_U^{gaze} is a free parameter for the degree to which maximum 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 maximum uncertainty in the first trial (that is, 0.8), to standardize the uncertainty-based weight to reflect the effect of maximum 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 model computes an RPE:

$$\delta_t^{\mathrm{RPE}} = \mathbf{R}_t - V_t(s) \quad (4)$$

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

$$V_{t+1}(s) = V_t(s) + \alpha \times \delta_t^{\mathrm{RPE}} \quad (5)$$

where $V_t(s)$ is the value of the CS s presented 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 hence we modelled the influence of model-free learning on the trial-by-trial gaze index in the model 'value' via:

$$\text{Model MF}: \mathbf{GazeIndex}_t = c + \beta_V^{\mathrm{gaze}} \times \mathbf{V}_t(s) \quad (6)$$

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. Lastly, we constructed a model assuming that 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:

$$\text{Model MF + MB}: \mathbf{GazeIndex}_t = c + \beta^{\mathrm{gaze}} \times \left([1-\omega] \times \mathbf{V}_t(s) + \omega \times \tilde{\mathbf{U}}_t(s)\right) \quad (7)$$

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 $\tilde{U}_t(s) = (U_t(s) - 0.4) \times \frac{2}{0.4}$. Note, that it is therefore difficult to interpret the absolute size of the weighting parameter ω, but that it is useful to analyse differences in parameter estimates between groups.

Parameter estimation and model comparison. To perform model comparison, we estimated free model parameters using maximum likelihood estimation for each individual participant, assuming Gaussian residuals. Bounded parameters were transformed to an unbounded scale for fitting: learning rate parameters for learning from RPEs or SPEs, 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 $\beta_U = 0.8 \times \exp(b_U)$. Optimization was performed using the nlm function in the R package stats[48]. We compared models for each participant 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: (1) a stronger value effect in sign-trackers (that is, stronger evidence for the model assuming conditioned responses to CS value[13,16,17]); and (2) a stronger model-based[2,3] uncertainty response in goal-trackers.

To increase stability in the estimation of noisy model parameters, we followed up maximum likelihood estimation 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 $\mu_\alpha = 0.3$ (that is, $\mu_a = \log[0.3/0.7]$) and a standard deviation of $\nu = 5$.

Eye-tracking analyses of pupil dilation. Pupil size data for valid fixation samples were z-standardized for each participant. For each participant and calibration (trials 1–8 and trials 9–16), we corrected for average baseline pupil size during 1 s 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 analysed pupil size via repeated-measures ANOVA with the factors trials (3–8 versus 9–16), CS value (€−2 to €+2) and time within trial (6 s 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 with 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 versus 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 6 s from CS onset to US onset. We thus tested whether the effects of CS value and of trials changed as a function of linear time within trials (seconds 1–6 after CS presentation (that is, the interaction trials × time and the interaction CS value × trials × time)). Next, we focused our analysis on the last second before US onset, where the signal is least confounded by luminance-related influences from CS presentation. Estimated contrasts were extracted from the ANOVA for visualization.

To visualize the effects of trials and of CS value, we moreover performed random-effects 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 participants with CS value effects deviating more than 6 s.d. from the mean, yielding a total of two excluded data points. We performed repeated-measures ANOVA on the estimated regression coefficients for (1) the intercept and (2) the linear CS value effect, with the 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 of: (1) the trial effect (3–8 versus 9–16) additionally nested within sign- versus goal-trackers; and (2) the difference between sign- and goal-trackers in the CS value effect, nested within trials (3–8 versus 9–16).

Pupil analyses using computational modelling. 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, when 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:

$$\mathbf{pupil}_t = c + \beta_V^{\mathrm{pupil}} \times \mathbf{V}_t(s) \quad (8)$$

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:

$$\mathbf{pupil}_t = c + \beta_U^{\mathrm{pupil}} \times \mathbf{U}_t(s) \quad (9)$$

Again, c is a constant, here capturing pupil size independent of learning. As for modelling gaze direction, we again performed 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 positive values using an exponential transform $\beta_U^{\mathrm{pupil}} = \exp(b_U^{\mathrm{pupil}})$. We used weakly informative Gaussian priors with a prior mean for the learning rate of $\mu_\alpha = 0.3$ (that is, $\mu_a = \log[0.3/0.7]$), a prior mean for the regression parameter of $\mu_{\beta t} = 0$, and prior standard deviations of $\nu = 5$. Due to the large noise in pupil size, we obtained fixed-effects 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 refitted to these normalized data for visualizing model predictions. The results from these analyses are shown in Fig. 2f.

Behavioural analyses. *Forced-choice task.* Data for the forced-choice task were available for 39 goal-tracker participants and 42 sign-tracker participants. Successful Pavlovian learning was assessed via the percentage of correct choices in the forced-choice task. We tested for a group difference in the percentage of correct choices.

Instrumental conditioning. Data on instrumental conditioning were available for all 43 sign-trackers and 43 goal-trackers. We measured overall learning speed as the number of trials needed until the learning criterion was reached (with a minimum of 60 and a maximum of 120 learning trials), and tested learning speed in sign-trackers versus goal-trackers. Moreover, to measure initial learning, we extracted the first 20 trials. To measure asymptotic learning, we extracted the last 20 trials. For these, we computed the difference in the response rates between instrumental conditions (collect versus leave) for each participant, and tested the effect in sign-trackers versus goal-trackers.

PIT. Data on the PIT task were available for all 43 sign-trackers and 41 goal-trackers. 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 participants via t-tests. We tested the strength of the PIT effect in sign-trackers and goal-trackers, and performed one-tailed tests of the a priori hypothesis[19] that PIT effects are stronger and more frequently individually significant in sign-trackers compared with goal-trackers.

fMRI analyses. *Preprocessing.* fMRI recordings were preprocessed using Nipype[61]. First, correction for differences in slice time acquisition to the middle slice was performed. Voxel-displacement maps were estimated based on the acquired field maps. All images were realigned to correct for head motion, distortion and their interaction. After coregistration of the individual structural T1 images to the individual mean EPI, the structural image was spatially normalized with a resampling resolution of $2\times2\times2\,mm^3$, 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. Before statistical analysis, data were high-pass filtered with a cut-off of 128 s.

Value learning during Pavlovian conditioning. We performed model-based fMRI analyses via first- and second-level analyses in SPM. We used the model-free reinforcement learning model (see equations (4) and (5)) to compute the trial-by-trial value of the CS V_t. Based on the reinforcement learning model, we determined the trial-by-trial temporal difference 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 (that is, $RPE_{US}=R_t-V_t(s)$).

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 (see Supplementary Information), but also indicated good robustness with respect to 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 3 s within one onset regressor. Stimulus onsets were parametrically modulated by the trial-by-trial temporal difference RPE. Additional nuisance regressors captured variance specific to US onsets, the eyetracker recalibration after trial 40, fixation reminders, and realignment parameters with derivatives[62]. Regressors were convolved with the canonical haemodynamic response function.

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 BOLD response in the current model, by including temporal and dispersion derivatives of the haemodynamic response function 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 sign- and goal-tracker groups.

Individual participants' parameter estimates for the RPE parametric modulator were taken to the second level. Valid fMRI recordings during Pavlovian conditioning were available for 39 sign-trackers and 39 goal-trackers. A two-sample t-test was performed comparing the RPE effect between sign-trackers and goal-trackers, with testing site as a control covariate of no interest. Differences between sign-trackers and goal-trackers 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 participants) – 3 (parameters used for the mean signals in sign-trackers and goal-trackers 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 RPE response in sign-trackers compared with goal-trackers. The visualization threshold was $P_{uncorrected}<0.005$ ($k=0$.) Statistical testing was performed in an a priori defined VOI in the bilateral NAc[10]: we chose a previously validated bilateral ventral striatal VOI a priori from the IBASPM 71 atlas, and derived this from the Wake Forest University PickAtlas software (www.fmri.wfubmc.edu/software/PickAtlas). RPEs in learning tasks akin to ours have been reported in this very VOI on numerous previous occasions (for example, refs. [63–66] 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[67]: 78% of our a priori VOIs were inside the VOI from the meta-analysis. Moreover, we performed a meta-analysis at https://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 FWE correction within the VOI to control for multiple comparisons.

Analyses of appetitive trials. While a wealth of evidence supports positively coded appetitive RPEs in striatal dopamine activity, the coding of aversive RPEs remains less clear. Some evidence suggests that aversive RPEs 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 and €+2). We coded the onsets of win-predictive and neutral CSs in one onset regressor, while the onsets of loss-predictive CSs were modelled as a separate onset regressor. The win- and neutral-predictive CSs were parametrically modulated by trial-by-trial temporal difference appetitive RPEs. 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-trackers 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 VTA[1] are known to project not only to the NAc, but also to the dorsal striatum (putamen and caudate), amygdala and vmPFC, and may drive fMRI BOLD correlates of RPE signals in these areas[68]. Sign-trackers do show increased CS-related activity in a range of different regions of the brain reward system[32], but whether these resemble RPEs is unclear. Here, we tested for RPE-like signals in several a priori VOIs thought to carry RPE-like BOLD responses, including the putamen, caudate, VTA, amygdala and vmPFC. VOIs were taken from ref. [69]. Results are reported for the average RPE signal in these VOIs. We first performed a priori tests using ANOVA with the factors group (sign- versus goal-trackers) and VOI. We did so for our a prior analysis involving gains and losses, and in addition for the analysis of gains only. We performed exploratory tests for each group of sign- and goal-trackers (one-tailed test of a signal larger than zero), and we performed one-tailed statistical tests of the hypothesis[32] that the RPE-like signals are stronger in sign-trackers than goal-trackers, which we also corrected for multiple exploratory tests. The visualizations of results from voxel-vise analyses are based on uncorrected thresholds of $P_{unc}<0.005$ ($k=40$), $P_{unc}<0.01$ ($k=40$; all VOIs) and $P_{unc}<0.5$ ($k=40$; VTA).

Exploratory analyses 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 $P<0.005$.

State learning during Pavlovian conditioning. The learning of model-based state transitions relies on SPEs (see 'Model-based influences on gaze'), which have previously been reported in the IPS and lPFC[4]. To estimate a neural SPE signal, we used the trial-by-trial SPE (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 lPFC 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 (ref. [70]) from the probabilistic brain atlas (Jülich-Düsseldorf cytoarchitectonic atlas) using the Anatomy Toolbox[71]. The lateral PFC VOI was extracted from the Wake Forest University 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-trackers than in sign-trackers, and whether there was an interaction of group × 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 visualized voxel-based results for the group difference based on uncorrected thresholds of $P_{unc} < 0.005$ ($k = 40$) and $P_{unc} < 0.01$ ($k = 40$).

Alternative classification of sign- and goal-trackers. *Using computational modelling 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'):

$$\mathbf{GazeIndex}_t = c + \beta_U^{\mathrm{gaze}} \times \mathbf{U}_t + \beta_V^{\mathrm{gaze}} \times \mathbf{V}_t(\mathbf{s}_t) \quad (10)$$

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 participant. As before, we performed 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 $\mu_a = 0.3$ (that is, $\mu_a = \log[0.3/0.7]$)) 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 participant to classify individuals as sign- or goal-trackers. The third of participants ($n = 43$) with the most positive weight parameter were classified as sign-trackers, whereas the third of participants ($n = 43$) with the most negative weight parameter were classified as goal-trackers.

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 to test for a larger PIT effect in sign-trackers. For the neural analyses, we tested whether sign-trackers showed an RPE signal averaged across all tested VOIs, and whether it was stronger than in goal-trackers. Likewise, for the neural SPE, we tested whether the difference between sign- and goal-trackers differed between VOIs (IPS and lPFC), whether the SPE signal was stronger in sign- than goal-trackers in IPS, and whether each group showed an 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:

$$P\left(T_i^t | \mathbf{CS}_i^t, \beta_i^t\right) = \frac{1}{B\left(\beta_i^t\right)} \prod_{j=1}^{J} \left(T_{i,j}^t\right)^{\beta_{i,j}^t - 1}$$

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 is 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, $\delta_t^{\mathrm{SPE}} = 1 - E\left[T_{i,j}^t\right]$.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are available for Figs. 1–6 and Supplementary Figs. 2–12. Data sharing will be based on acceptance by the study team that: (1) a valid and timely scientific question, based on a written protocol, has been posed by those seeking to access the data; and (2) 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.

Code availability

Experimental code is freely available on request to the corresponding author. Analysis code will be provided with data access.

Received: 20 October 2017; Accepted: 25 September 2019;
Published online: 11 November 2019

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. in *Stevens' Handbook of Experimental Psychology* 3rd edn 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**, 587 (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**, 869–873 (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* **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**, 1936–1947 (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.* **269**, 295–308 (2019).
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**, 1428–1441 (2013).
43. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
44. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
45. Garbusow, M. et al. Pavlovian-to-instrumental transfer in alcohol dependence: a pilot study. *Neuropsychobiology* **70**, 111–121 (2014).
46. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV* (American Psychiatric Publishing, 1994).
47. 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).
48. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2016).
49. Singmann, H., Bolker, B., Westfall, J. & Aust, F. afex: Analysis of Factorial Experiments R package version 0.18-0 https://cran.r-project.org/web/packages/afex/index.html (2017).
50. Lenth, R. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.1. https://cran.r-project.org/web/packages/emmeans/index.html (2018).
51. 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**, 688–690 (2006).
52. Canty, A. & Ripley, B. D. boot: Bootstrap R (S-Plus) functions. R package version 1.3-18. https://cran.r-project.org/web/packages/boot/ (2017).
53. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Applications* (Cambridge Univ. Press, 1997).
54. Morey, R. D. Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* **4**, 81–84 (2008).
55. Kelley, K. MBESS: The MBESS R Package. *R version 4.5.1.* https://cran.r-project.org/web/packages/MBESS/index.html (2019).
56. Hogarth, L., Dickinson, A. & Duka, T. in *Attention and Associative Learning: From Brain to Behaviour* (eds Mitchell, C. J. & Le Pelley, M. E.) 71–98 (Oxford Univ. Press, 2010).
57. 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**, 11182–11191 (2009).
58. Hickey, C., Chelazzi, L. & Theeuwes, J. Reward changes salience in human vision via the anterior cingulate. *J. Neurosci.* **30**, 11096–11103 (2010).
59. Hickey, C. & van Zoest, W. Reward creates oculomotor salience. *Curr. Biol.* **22**, R219–R220 (2012).
60. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**, 194–203 (2001).
61. Gorgolewski, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* **5**, 13 (2011).
62. Iglesias, S. et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80**, 519–530 (2013).
63. Deserno, L. et al. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl Acad. Sci. USA* **112**, 1595–1600 (2015).
64. 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).
65. 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**, 4487–4493 (2013).
66. Gluth, S., Hotaling, J. M. & Rieskamp, J. The attraction effect modulates reward prediction errors and intertemporal choices. *J. Neurosci.* **37**, 371–382 (2017).
67. Garrison, J., Erdeniz, B. & Done, J. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 1297–1310 (2013).
68. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
69. Nebe, S. et al. No association of goal-directed and habitual control with alcohol consumption in young adults. *Addict. Biol.* **23**, 379–393 (2018).
70. Neyens, V. et al. Representation of semantic similarity in the left intraparietal sulcus: functional magnetic resonance imaging evidence. *Front. Hum. Neurosci.* **11**, 402 (2017).
71. Eickhoff, S. B. et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–1335 (2005).

Acknowledgements

This work was supported by the German Research Foundation (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, and in part by CRC-TR 265). E.F. is a participant in the BIH Charité Clinician Scientist Program funded by the Charité—Universitätsmedizin Berlin and Berlin Institute of Health. Q.J.M.H. acknowledges support from the UCLH NIHR BRC. S.N. received funding from the University of Zurich Grants Office (FK-19-020). We thank N. B. Krömer for helpful feedback and advice on the analyses, S. Kuitunen-Paul for helpful feedback, and M. Rothkirch for help with setting up eye-tracking at the Berlin site. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Q.J.M.H. conceived of the study idea. M.A.R., E.F., H.-U.W., U.S.Z., H.W., P.S., M.N.S., F.S., A.H. and Q.J.M.H. designed the study. D.J.S., M.G., M.S., S.N., E.O., E.F., U.S.Z., M.N.S., F.S., A.H. and Q.J.M.H. conducted the implementation, pilots and setup. M.G., M.S., S.N. and C.S. acquired the data, with supervision from N.R.-S., H.-U.W., U.S.Z., H.W., P.S., M.N.S., F.S., A.H. and Q.J.M.H. D.J.S. analysed the data, with supervision from M.A.R., P.D. and Q.J.M.H. and input from L.D., M.R., F.S. and A.H. D.J.S., M.A.R., P.D. and Q.J.M.H. wrote the manuscript. All authors read and revised the manuscript and provided critical intellectual contributions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0765-5.

Correspondence and requests for materials should be addressed to D.J.S.

Peer review information Primary Handling Editor: Marike Schiffer.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Daniel J. Schad

Last updated by author(s): Sep 1, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The task was programmed using Matlab 2011 (MATLAB version 7.12.0, 2011; MathWorks, Natick, MA, USA) with the Psychophysics Toolbox Version 3 extension.
Data analysis	Data were analyzed using Matlab 2013a (MATLAB version 8.1.0.604, 2013; MathWorks, Natick, MA, USA) and the R System for Statistical Computing Version 3.3.2 (http://www.r-project.org). fMRI data were analyzed using Nipype (for preprocessing; Gorgolewski et al., 2011) and Statistical Parametric Mapping 8 (SPM8; http://www.fil.ion.ucl.ac.uk/spm/; Wellcome Department of Imaging Neuroscience).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Data are quantitative experimental.
Research sample	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 (LeAD study; www.lead-studie.de; clinical trial number: NCT01679145). We therefore drew healthy male participants with an age of 18 years from the general population in two testing sites (Berlin and Dresden, Germany).
Sampling strategy	We randomly drew healthy male participants from the general population via local registries. The study was designed and sample size of 198 subjects was chosen to detect moderate differences (a priori Power analysis: d=0.4, α=0.05, ß=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, α=0.05, ß=0.79).
Data collection	Data were collected using computers, manual response buttons, an eye-tracker, and fMRI recordings. Only the participant and the researcher were present during data collection. Occasionally, multiple researchers were present during the testing session. The researchers were blind to group definitions (sign- versus goal-trackers) during the data collection.
Timing	Data collection started on Feburary 19, 2013, and stopped on October 31, 2015.
Data exclusions	A priori exclusion criteria were left-handedness, a history of any substance dependence or current substance use except for nicotine dependence, other major psychiatric disorders (DSM-IV axis I; CIDI) and neurologic disorders. 198 subjects were tested in total. Valid eye-tracking data was available for 129 subjects, from which 43 sign-trackers and 43 goal-trackers could be identified (for description of the statistical selection approach see the manuscript). For each task or data source (behavior; pupillometry; fMRI), data from some subjects was missing for technical reasons (e.g., early abortion of task execution, recording problems). The final sample sizes per task and recording technique are indicated in the methods section.
Non-participation	Invited via letter (addresses from resident registration offices): n = 1937; Interested persons: n = 475; Screened: n = 445; Included: n = 244; Valid assessment + MRI (available at time of analysis): n = 198; Valid eye-tracking data: n = 129; Classified as sign-/goal-trackers: n = 43/43; Further missing data for each task or measure, see Methods section.
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.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	Participants were drawn from the general population via local registries; participants were contacted via postal letters concerning study participation. It is not clear how potential self-selection biases should impact the results as we defined groups based on subtle experimental eye-tracking markers.
Ethics oversight	Ethical approval for the study was obtained from the ethics committee of Charité-Universitätsmedizin Berlin (EA1/157/11) and Universitätsklinikum Dresden (EK228072012).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	task; event-related
Design specifications	In the fMRI scanner, subjects performed Pavlovian conditioning and Pavlovian-instrumental transfer (PIT). Pavlovian conditioning: 80 trials per subject; each trial lasted 9 sec; the inter-trial-interval was exponentially distributed with a minimum of 2 sec and a maximum of 6 sec (mean duration: 2.9 sec). PIT: 90 trials involving conditioned CSs. Data from additional 72 trials involving drink-related background stimuli will be reported elsewhere. Each trial involved an initial 600 ms of Pavlovian CS presentation, an instrumental response window of 3 sec, 300 ms visualization of response results, and an exponentially distributed inter-trial-interval (min / mean / max duration = 2 / 2.9 / 6 sec).
Behavioral performance measures	During Pavlovian conditioning attention was monitored using eye-tracking and pupillometry. We studied probabilities (mean + SD) of fixating a presented CS, the background, and the location of later US presentation, and how this was affected by time and CS value. We also studied baseline-corrected pupil size. Pavlovian learning was moreover measured after the fMRI scanning session in a forced choice task via the percentage (mean + SD) of correctly choosing the higher-valued out of two presented CSs. During the Pavlovian-instrumental transfer (PIT) task conducted during fMRI scanning, the number of button presses per trial was recorded. Responses were made on a 1 x 4 current design MR-compatible response box button using the dominant index finger. We analyzed response rate (mean + SD) as a function of CS value and performed linear regression of the number of button presses on CS value (mean + SD of regression coefficients). SD: standard deviation across subjects.

Acquisition

Imaging type(s)	functional MRI
Field strength	3
Sequence & imaging parameters	Pulse sequence type: gradient echo; Imaging type: EPI; Field of view: 192 x 192 mm^2; Matrix size: 64 x 64 x 42; Slice thickness: 2 mm, 1mm gap between slices; Orientation: approximately -25° to the bicommissural plane; TE / TR / Flip angle: 25 ms / 2410 ms / 80°
Area of acquisition	whole brain scan

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Nipype (release 0.11.0; hash of git version: 07d598aaffce0ecd10a3bfbed63369a6634259e5) (Gorgolewski et al., 2011) using SPM8; slice time correction with ref. to middle slice, motion correction via realignment to 1st slice of each volume, correction for field inhomogeneity with voxel displacement maps, co-registration ind. mean EPI on ind. T1 image, segmentation of ind. T1 image , normalization, smoothing (8mm FWHM Gaussian kernel)
Normalization	segmentation and normalization of ind. T1 images via default ICBM (European Brains) template (MNI space), application of normalization parameters to ind. distortion-corrected EPI images, resampling to 2x2x2mm^3 voxel size
Normalization template	default template in SPM8/New Segmentation (ICBM152 - European Brains)
Noise and artifact removal	inclusion of volume-to-volume realignment parameters (3 translation, 3 rotation) on individual-statistics level; no removal of tissue or physiological signals/artifacts
Volume censoring	no censoring used

Statistical modeling & inference

Model type and settings	mass univariate first-level: random effects model, high-frequency pass filter: 128s second-level: random effects t-test
Effect(s) tested	Model-based fMRI: one regressor coding all 2 (stimuli per trial: CSs + USs) x 80 (trials) = 160 stimulus onsets; the regressor of key interest was a parametric modulator coding the temporal difference reward prediction error obtained from a reinforcement learning model for each stimulus and trial; an additional model parametrically modulated 80 onsets coding a state prediction error signal during US onset in each trial; we tested these parametric modulators in sign- and goal-tracker groups as well as the group difference via a 2nd-level t-test; t-contrasts were used to test the effect in each group; the group difference was tested via an F-contrast

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s)	We used a NAc ROI derived from the Wake Forest University (WFU) PickAtlas software (www.fmri.wfubmc.edu/software/PickAtlas). Additional ROIs for analysis of reward prediction errors were taken from a recent paper by Nebe et al. (2018). ROIs for state prediction errors were taken from Neyens et al. (2018; IPS) and from the WFU PickAtlas software (IPFC).
Statistic type for inference (See Eklund et al. 2016)	FWE corrected voxel-wise threshold of p<.05 (ROI-based), and mean signal per ROI
Correction	FWE correction in volume of ROI mask

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis