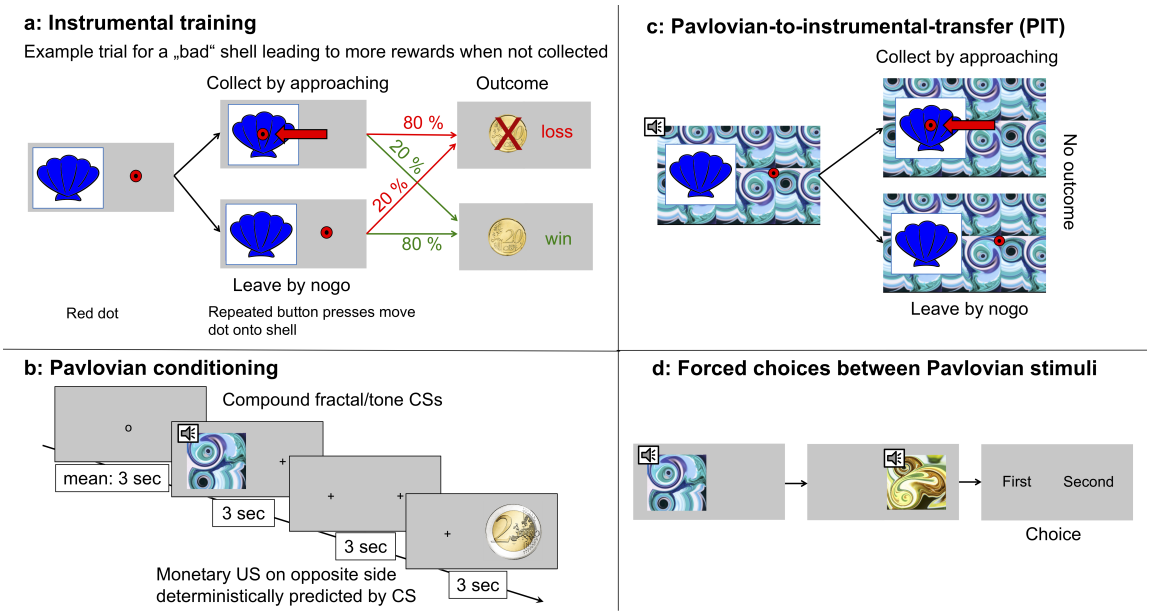
Supplementary Information

1 Supplementary Methods



Supplementary Figure 1 | **The Pavlovian-instrumental transfer (PIT) paradigm** consisted of four parts. **a.** Instrumental training. Subjects learned to collect good shells via repeated button presses and to avoid bad shells by doing nothing. **b.** Pavlovian conditioning: CSs were followed by monetary wins or losses ($-2, -1, 0, +1, +2$ €). Gaze position was recorded via eye-tracking. **c.** PIT. Subjects performed the instrumental task in nominal extinction (i.e., without presentation of outcomes). The background was tiled with Pavlovian CSs. **d.** Forced choices. Subjects were asked to choose the higher-valued out of two CSs.

1.1 Learning from wins vs. losses

Control analyses examined differences in learning from wins and losses. The neural basis of model-free RPE has been found to show differences between learning from wins versus losses [1], and our fMRI analyses of RPE supported this dissociation (c.f. Supplementary Figure 7). We therefore tested

whether wins and losses also differed in the learning process itself. To this end, we implemented computational model-free reinforcement learning models allowing for differences in the speed of learning (i.e., the reward learning rate parameter) or in the weight of CS value (i.e., the weight parameter) between wins and losses, to test whether assuming differences in these parameters provided a better account of the gaze direction or the pupil dilation data. We used the standard model-free reinforcement learning model (see equations (4) and (5)) as a baseline and compared it to two more complex models. First we compared it to a model allowing for two different learning rates for wins versus losses. Here, the update equation (equation 5) for $R_t > 0$ was:

$$V_{t+1}(s) = V_t(s) + \alpha_{win} \cdot \delta_t^{RPE} \ , \tag{11}$$

and for $R_t < 0$, it was:

$$V_{t+1}(s) = V_t(s) + \alpha_{loss} \cdot \delta_t^{RPE} \ , \tag{12}$$

where α_{win} indicates a free learning rate parameter for wins, and α_{loss} indicates a free learning rate parameter for losses. We fitted this model to the gaze index and to the pupil dilation data, and performed model comparison using BIC values to compare the model involving two learning rates to the simpler RL model using the same learning rate for wins and losses.

Next, we used a model allowing for two different weight parameters for wins versus losses. For $R_t > 0$ gaze direction was influenced by CS value via

$$\text{GazeIndex}_t = c + \beta_{V,win}^{\text{gaze}} \cdot V_t(s_t) \ , \tag{13}$$

and for $R_t < 0$, it was influenced via

$$\text{GazeIndex}_t = c + \beta_{V,loss}^{\text{gaze}} \cdot V_t(s_t) \ , \tag{14}$$

where $\beta_{V,win}^{\text{gaze}}$ was a free parameter indicating the CS value weight for wins, and $\beta_{V,loss}^{\text{gaze}}$ was a free parameter indicating the CS value weight for losses. We also defined the corresponding model assuming influences of CS value on pupil dilation, where for $R_t > 0$ pupil size was influenced by CS value via

$$\text{pupil}_t = c + \beta_{V,win}^{\text{pupil}} \cdot V_t(s_t) \ , \tag{15}$$

and for $R_t < 0$, it was influenced via

$$\text{pupil}_t = c + \beta_{V,loss}^{\text{pupil}} \cdot V_t(s_t) \ . \tag{16}$$

Again, $\beta_{V,win}^{\text{pupil}}$ was a free parameter indicating the CS value weight for wins, and $\beta_{V,loss}^{\text{pupil}}$ was a free parameter indicating the CS value weight for losses.

Model parameters were estimated via maximum likelihood estimation (MLE) and BIC values were computed. Random effects analysis was used for the gaze direction data, whereas the noisy pupil data was analyzed via fixed effects estimation.

1.2 Instructions for the PIT phase

1.2.1 Original german instructions

- Ok, jetzt kommt der wichtigste Teil. Sie sind nun wieder beim Muscheln Sammeln.

- Sammeln Sie weiter diejenigen Muscheln ein, die gut waren und lassen Sie weitherhin die schlechten liegen - dafür bekommen Sie weiterhin je 20 Cent.
- Wenn Sie schlechte Muscheln einsammeln, oder gute liegen lassen, verlieren Sie weiterhin 20 Cent.
- Im Hintergrund sehen Sie nun die bunten Bilder.
- Sie haben gelernt, dass jedem Bild ein bestimmter Geldbetrag zugeordnet ist. Dieser Wert wird Ihnen jetzt zufällig in der Hälfte der Fälle gutgeschrieben oder abgezogen.
- Es geht also immer um den Betrag, der mit dem Bild verbunden ist. Wenn also im Hintergrund das Bild ist, das mit $+1$ Euro verbunden ist, dann können Sie in der Hälfte der Fälle 1 Euro gewinnen. Wenn im Hintergrund das Bild ist, dem ein Verlust von 1 Euro zugeordnet ist, dann verlieren Sie in der Hälfte der Fälle 1 Euro.
- Das erfolgt aber ohne Ihr Zutun automatisch durch den Computer.
- Konzentrieren Sie sich also darauf, durch richtiges Muschelsammeln so viel Geld zu gewinnen wie möglich.
- Wieviel Sie in jeder Runde gewinnen oder verlieren, sei es wegen dem Hintergrundbild, oder wegen der Muschel, wird nicht mehr angezeigt. Ihr Gesamtguthaben am Ende des Experiments wird Ihnen ausgezahlt.
- Sammeln Sie also einfach weiter gute Muscheln ein.

1.2.2 English translation

- Okay, here's the most important part. You are now again collecting shells.
- Collect those shells that were good and leave the bad ones - you will still get 20 cents each.
- If you collect bad shells or leave good ones, you will continue to lose 20 cents.
- In the background you can now see the colorful pictures.
- You have learned that a certain amount of money is associated with each picture. This value is now credited or deducted to you by chance in half of the cases.
- So it is always about the amount that is associated with the picture. So if in the background there is the picture associated with $+1$ Euro, then you can win 1 Euro in half of the cases. If in the background is the image with which a loss of 1 Euro is associated, then you lose 1 Euro in half of the cases.
- However, this is done automatically by the computer without your contribution.
- So concentrate on winning as much money as possible by collecting shells properly.
- How much you win or lose in each round, whether because of the background image or because of the shell, is no longer displayed. Your total balance at the end of the experiment will be paid to you.
- So just keep collecting good shells.

1.3 Reporting policies

1.3.1 Replication

Findings on eye-movements and PIT replicate previous human results [2] and neural findings replicate previous animal results [3]. We triangulate results across three different experimental tasks, across three recording techniques (eye-tracking - gaze and pupil size - , behavioural responses, fMRI), and across different alternative group definitions.

1.3.2 Blinding of experimental group assignment

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

2 Supplementary Results

2.1 Eye-tracking

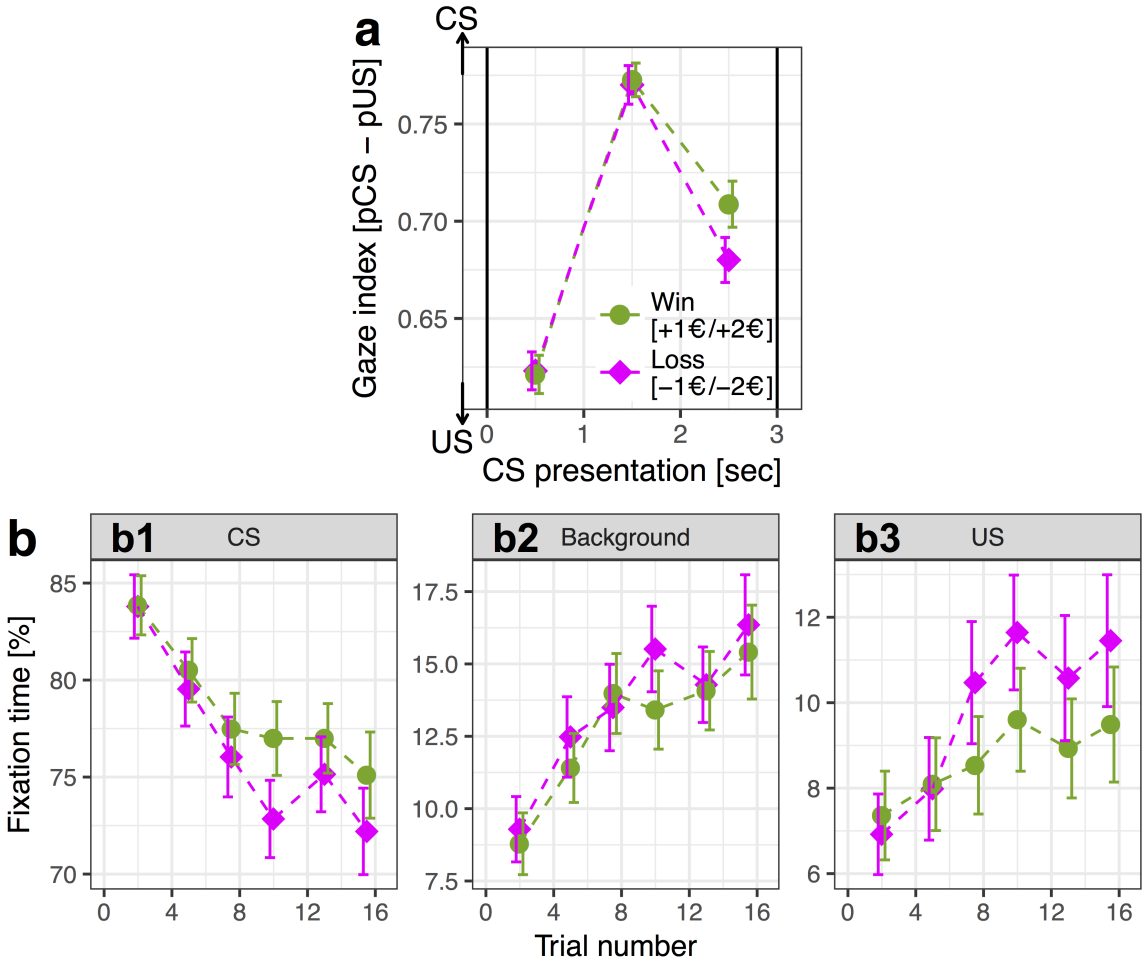
2.1.1 Effects of trial number

In the last second of CS presentation, gaze progressively shifted from the CS to the US location and the background over trials. Specifically, there was a three-way interaction between *trial number*, *location* (CS, US-location, background) and *time of CS presentation* (seconds 1, 2, and 3), $F(2.7, 347.5) = 8.58, p < .0001, \eta_p^2 = 0.004, 95\%$ confidence intervals $CI = [0.001\ 0.008]$ (Mauchly's test revealed a violation of the sphericity assumption, $W = 0.42, p < .001$, which was corrected via the Greenhouse-Geisser estimate of sphericity, $\epsilon = 0.68$). Post-hoc testing showed that the interaction *trial number* x *fixation location* was strongest for the last second of CS presentation, $F(2, 410) = 37.93, p < .0001, \eta_p^2 = 0.013, 95\%\ CI = [0.007\ 0.019]$, and was relatively weaker, yet still strong, in second 2, $F(2, 410) = 22.55, p < .0001, \eta_p^2 = 0.008, 95\%\ CI = [0.004\ 0.013]$, and in second 1, $F(2, 410) = 9.57, p = .0001, \eta_p^2 = 0.003, 95\%\ CI = [0.0009\ 0.007]$. For the last second of CS presentation (see Supplementary Fig. 2b), percentage fixation times decreased across trials for the CS, $t_{410} = -8.62, p < .0001, b = -0.008, SE = 0.001, 95\%\ CI = [-0.010\ -0.007]$, but increased for the US location, $t_{410} = 3.21, p = .001, b = 0.003, SE = 0.001, 95\%\ CI = [0.001\ 0.005]$, and for the background, $t_{410} = 5.41, p < .0001, b = 0.005, SE = 0.001, 95\%\ CI = [0.0033\ 0.0072]$.

2.1.2 Effects of CS value

The value of the CSs influenced gaze location. There was a three-way interaction between *CS value* (-2, -1, 0, +1, +2 €), *fixation location* (CS, US-location, background) and *time of CS presentation* (seconds 1, 2, and 3), $F(11, 1367) = 3.04, p = .0006, \eta_p^2 = 0.004, 95\%\ CI = [0.001\ 0.008]$ (the violation of the sphericity assumption, $W = 0.013, p < .001$, was corrected via $\epsilon = 0.67$). Post-hoc testing revealed the interaction *CS value* x *fixation location* was absent during the first two seconds of CS presentation (sec 1: $p = .882$; sec 2: $p = .059$, Holm-corrected for three seconds; see Supplementary Fig. 2a). However, the interaction was significant during the third second of CS presentation (see Supplementary Fig. 2a+b), $F(8, 2267) = 4.46, p_{Holm} < .0001, \eta_p^2 = 0.005, 95\%\ CI = [0.002\ 0.009]$. Planned contrasts testing a linear effect of CS value (-2, -1, 0, +1, +2 €) showed that percentage fixation time increased with increasing CS value on the CS ($t_{2267} = 3.48, p = .0005, b = 0.061, SE = 0.018, 95\%\ CI = [0.027\ 0.096]$), but marginally decreased with increasing CS value on the US-location ($t_{2267} = -1.78, p = .075, b = -0.031, SE = 0.018, 95\%\ CI = [-0.066\ 0.003]$) and on the background ($t_{2267} = -1.70, p = .089, b = -0.030, SE = 0.018, 95\%\ CI = [-0.065\ 0.005]$). Thus, during the third second of CS presentation CSs signaling wins became attractive to the gaze, while CSs signaling loss came to repel the gaze (see Fig. 1c and Supplementary Fig. 2a+b).

We tested how the gaze response to CS value developed across trials. There was a significant three-way interaction *CS value* x *trial number* x *location*, $F(5.8, 743.0) = 4.59, p = .0002, \eta_p^2 = 0.004, 95\%\ CI = [0.001\ 0.008]$ (violation of sphericity $W = 0.23, p < .001$ was corrected via $\epsilon = 0.73$) although the four-way interaction *CS value* x *trial number* x *location* x *time* failed to reach significance ($p = .192$). Tests of contrasts showed that the linear CS value effect increased across trials on the CS ($t_{1024} = 2.68, p = .007, b = 0.007, SE = 0.002, 95\%\ CI = [0.002\ 0.011]$) and on the US location (inverse CS value effect; $t_{1024} = -2.55, p = .011, b = -0.006, SE = 0.002, 95\%\ CI = [-0.011\ -0.001]$), but not on the background ($p = .897$). Moreover, we also tested the three-way interaction *CS value* x *trial number* x *location* for gaze during the third second of CS presentation



Supplementary Figure 2 | **A conditioned response in eye-gaze during Pavlovian conditioning. a.** Gaze direction index, i.e., the difference in the probabilities to fixate the CS minus the US-location, is displayed for CSs predicting win (green points) or loss (magenta diamonds) for the three seconds of CS presentation. **b1-3.** During the third second of CS presentation: Percentage valid fixation time displayed separately for the CS (**b1**), the location of later US-presentation (**b3**), and the background (i.e., the rest of the screen; **b2**) for CSs predicting wins and losses across trials. **a+b.** Error bars are SEM.

only, and found a significant interaction, $F(8.0, 2674.9) = 2.94, p = .003, \eta_p^2 = 0.003, 95\%\ CI = [0.0005\ 0.007]$. Here, the linear CS value effect increased linearly across trials for fixations on the CS ($t_{2674.9} = 2.47, p = .014, b = 0.008, SE = 0.003, 95\%\ CI = [0.002\ 0.015]$) and marginally on the US-location ($t_{2674.9} = -1.77, p = .077, b = -0.006, SE = 0.003, 95\%\ CI = [-0.013\ 0.0006]$), but not on the background ($p = .485$). We performed a test contrasting the increase in the linear CS value effect across trials for the CS (where across trials win-predictive CSs attracted more fixation time) with that at the US-location (where across trials win-predictive CSs attracted less fixation time), and found the 3-way interaction *linear CS value* x *linear trial number* x *fixation at CS minus US-location* to be significant ($t_{2674.9} = 2.45, p = .014, b = 0.014, SE = 0.006, 95\%\ CI = [0.003\ 0.026]$).

2.1.3 Gaze index

Based on prior research [3, 4], and to capture the above effects parametrically, we computed a gaze index as the difference between CS minus US fixation probability during the third second of CS presentation (Fig. 1b). Overall, gaze direction shifted from the CS towards the US and the background across trials ($p_{bootstrap} < .001, b = -0.011, SD_{subjects} = 0.024, SE = 0.002, 99.9\%\ CI = [-0.017\ -0.003]$), and gaze direction was biased towards the CS with higher CS value ($p_{bootstrap} < .05, b = 0.009, SD_{subjects} = 0.057, SE = 0.005, 95\%\ CI = [0.001\ 0.022]$). This latter effect increased across trials (*linear CS value* x *linear trial number*: $p_{bootstrap} < .05, b = 0.0015, SD_{subjects} = 0.0091, SE = 0.0008, 95\%\ CI = [0.00001\ 0.0032]$; Fig. 1c), suggesting a learning process.

2.2 Distinguishing sign- and goal-trackers

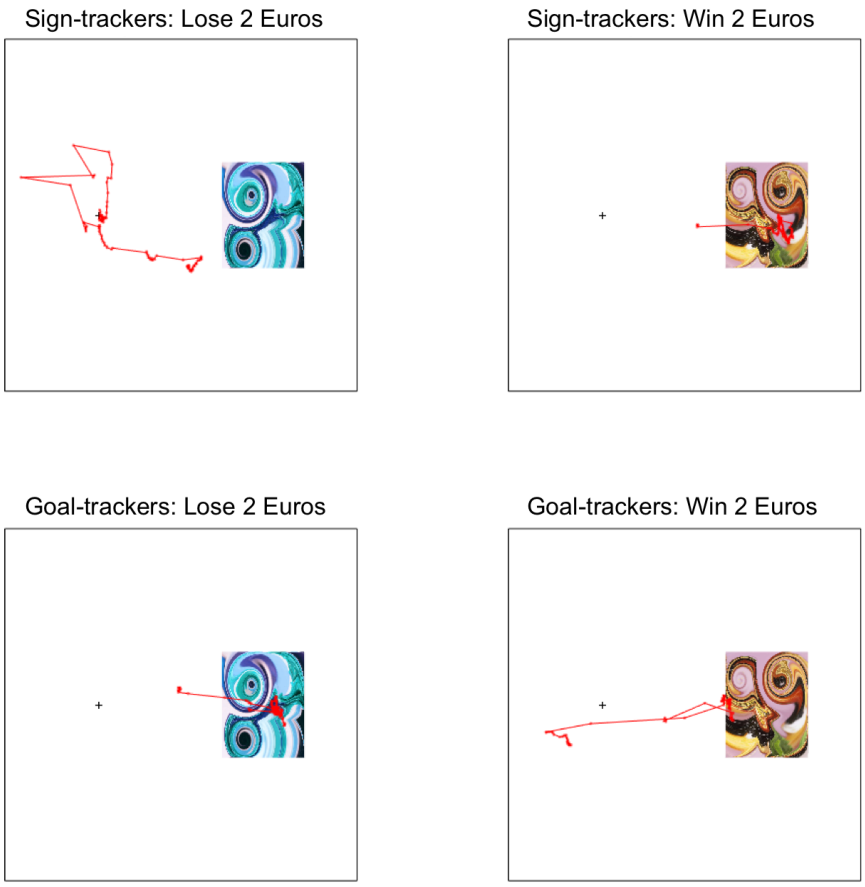
The histogram of CS value effects (regression coefficients) on gaze direction during the third second of CS presentation was used to assess ST and GT (Fig. 1c). The frequency of ST vs. GT did not significantly differ between testing sites (45.7% ST/GT in Berlin vs. 52.9% ST/GT in Dresden; $\chi^2(1) = 0.19, p = .7$), precluding substantial experimenter effects.

During the third second of CS presentation, ST gaze approached win-predictive CSs more than loss-predictive CSs, thus tracked the sign based on its value ($p_{bootstrap} < .001, b = 0.059, SD_{subjects} = 0.069, SE = 0.011, 99.9\%\ CI = [0.039\ 0.129]$, see Fig. 1d-f). ST showed continuous [5–7] signatures of learning to avoid the aversive CSs over time ($p_{bootstrap} < .001, b = -0.013, SD_{subjects} = 0.033, SE = 0.005, 99.9\%\ CI = [-0.037\ -0.001]$), without changing their responses to appetitive CSs ($p > .1$; trial number x CS value: $p_{bootstrap} < .05, b = 0.003, SD_{subjects} = 0.012, SE = 0.002, 95\%\ CI = [0.0003\ Inf]$; Fig. 1d+f). By contrast, GT approached the US location for predicted wins more than for predicted losses, thus tracked the goal based on its value ($p_{bootstrap} < .001, b = -0.036, SD_{subjects} = 0.024, SE = 0.004, 99.9\%\ CI = [-0.053\ -0.027]$; Fig. 1d-f). This inverse CS value effect did not increase over time (trial number x CS value: $p = .601$).

Supplementary Figure 3 displays gaze fixation location in representative trials during the first three seconds per trial after learning for sign- and for goal-trackers.

2.2.1 Robustness to outliers

The results were stable when excluding four formal outliers in the effect of CS value on the gaze index (deviating more than two standard deviations from the mean). After outlier-removal, for gaze direction the interaction *CS value* x *second of CS presentation* remained significant ($p = .007$), the CS value effect during second three was marginal (under one-tailed testing [4, 8, 9]: $p = .092$),



Supplementary Figure 3 | **Examples of gaze fixation locations for sign- and goal-trackers in aversive and appetitive Pavlovian trials.** The colourful pictures are Pavlovian CSs, presented during the first three seconds per trial. Fixation crosses are shown at the location of later US presentation. Gaze fixation location during these first three seconds of CS presentation is displayed in red. Upper panels show representative trials after learning from sign-trackers, and lower panels show representative trials after learning from goal-trackers. Left panels show loss-predictive CSs, right panels show win-predictive CSs.

and was stronger compared to seconds one or two ($p = .057/.002$). Moreover, the CS value effects in ST and in GT were reliable ($p_{bootstrap} < .001$).

2.2.2 Computational models of uncertainty and CS value

We tested whether model-based and model-free learning differentially explained gaze patterns in sign- and goal-trackers. We found that goal-trackers' gaze was best explained by the model assuming uncertainty-based gaze control ($t_{84} = -3.28, p = .002, \Delta$ BIC $= -3.73, SE = 1.14, 95\%\ CI = [-6.00\ \ -1.47]$; see Fig. 1g; after outlier exclusion: $p < .001$), supporting uncertainty-based attention in these individuals. In sign-trackers, to the contrary, the difference in BIC values between models was significantly shifted towards value-based gaze control; $F(1, 84) = 6.87, p = .010, \eta_p^2 = 0.038, 95\%\ CI = [0.002\ 0.110]$; after outlier removal: $p = .015$). While none of the models significantly outperformed the other ($p = .672$), numerical differences showed a slight advantage for gaze control by CS value, i.e., Pavlovian value-based attention.

Across all subjects, model parameters (fixed-effects for more stable estimates) showed relatively slow model-free reward learning with a learning rate of $\alpha = 0.173$ ($a = -1.563$), with large posterior uncertainty of $\sigma_a = 1.616$ [10]. The learning rate estimate for model-based state learning was of similar size with $\eta = 0.186$ ($n = -1.477$), but with much smaller posterior uncertainty of $\sigma_n = 0.219$.

We next fitted a dual-control model in which model-free value and model-based uncertainty were combined via a weighting parameter ω to determine attention: values of $\omega = 1$ indicated attention was guided purely by model-based uncertainty, whereas values of $\omega = 0$ indicated attention was guided purely by model-free value. The results on the weighting parameter (Fig. 1h) showed that goal-trackers strongly relied on model-based uncertainty ($M = 0.84, SD = 0.18$), whereas contributions from both systems were combined to guide gaze in sign-trackers ($M = 0.48, SD = 0.24$), reflecting a significant shift towards more model-free control ($p_{bootstrap} < .001, b = 0.36, SE = 0.05, 99.9\%\ CI = [0.20\ 0.50]$).

2.3 Pupil size

2.3.1 Statistical analysis

For pupil size, we first analyzed all data from CS onset to US onset, and then focussed analysis on the region of interest in the last second before US presentation, where we expect the signal to be least influenced by luminance-related confounds. In the overall analyses, we found a significant three-way interaction *trials* (3-8 vs. 9-16) x *time* (linear effect of seconds 1-6 of CS presentation) x *group* (sign- vs. goal-trackers): $t_{410} = 2.97, p = .003, b = 0.12, SE = 0.04, 95\%\ CI = [0.04\ 0.19]$. This was due to a significant interaction *trials* x *time* in goal-trackers ($t_{410} = -5.01, p < .0001, b = -0.270, SE = 0.054, 95\%/; CI = [-0.376\ \ -0.164]$), which was absent in sign-trackers ($p = .461$). In goal-trackers, pupil size significantly decreased across *trials* (3-8 vs. 9-16) for the last second before US presentation (i.e., second 6 after CS onset: $t_{140} = -2.29, p = .023, b = -0.055, SE = 0.024, 95\%\ CI = [-0.102\ \ -0.008]$; Fig. 2a) and for the second last second (i.e., second 5: $t_{140} = -2.56, p = .012, b = -0.061, SE = 0.024, 95\%\ CI = [-0.108\ \ -0.014]$), but did not change as a function of trials for seconds 1 to 4 ($p - values \geq .103$). Sign-trackers, to the contrary, did not show a change in average pupil size across trials, neither for the last second before US presentation (sec 6: $p = .405$, trials x group: $t_{140} = 2.20, p = .030, b = 0.04, SE = 0.02, 95\%\ CI = [0.004\ 0.071]$; sec 5: $p = .524$), nor averaged across all six seconds within a trial ($p = .207$).

Concerning influences from CS value on pupil size (see Fig. 2b), we found a significant four-way interaction *CS value* (linear effect: $-2, -1, 0, +1, +2$ €) x *trials* (3-8 vs. 9-16) x *time* (linear effect of seconds 1-6 of CS presentation) x *group* (sign- vs. goal-trackers): $t_{1640} = 2.69, p = .007, b = 0.542, SE = 0.201, 95\%\ CI = [0.147\ 0.937]$. Post-hoc contrasts showed the three-way interaction *CS value* x *trials* x *group* was significant in the last two seconds before US presentation (second 6: $t_{638} = 3.52, p = .0005, b = 0.285, SE = 0.081, 95\%\ CI = [0.126\ 0.444]$; second 5: $t_{638} = 2.78, p = .006, b = 0.225, SE = 0.081, 95\%\ CI = [0.066\ 0.384]$), but was not significant in the first four seconds after CS presentation ($p - values \geq .060$; second 1: $p = .133$). In the last second before US presentation, sign-trackers showed a significant interaction *CS value* x *trials* ($t_{638} = 2.93, p = .004, b = 0.339, SE = 0.116, 95\%\ CI = [0.112\ 0.566]$), which reflected an increase of the CS value effect across trials. Specifically, in sign-trackers pupil size increased with increasing CS value (linear effect) at the end of learning (trials 9-16; $t_{1314} = 2.89, p = .004, b = 0.521, SE = 0.180, 95\%\ CI = [0.167\ 0.874]$), but not at the beginning of learning (trials 3-8: $p = .383$). Moreover, in sign-trackers in trials 9-16, pupil size increased with increasing CS value (linear effect) in the last second before US presentation, but not in the previous seconds ($p - values \geq .288$; second 5: $p = .091$). In goal-trackers, to the contrary, CS value did not significantly increase or decrease pupil size, neither in trials 9-16 (second 6: $p = .112$, seconds 1 to 5: $p - values \geq .133$; averaged across seconds 1-6: $p = .258$) nor in trials 3-8 (for seconds 1-6: $p \geq .268$; averaged: $p = .495$).

2.3.2 Computational modeling

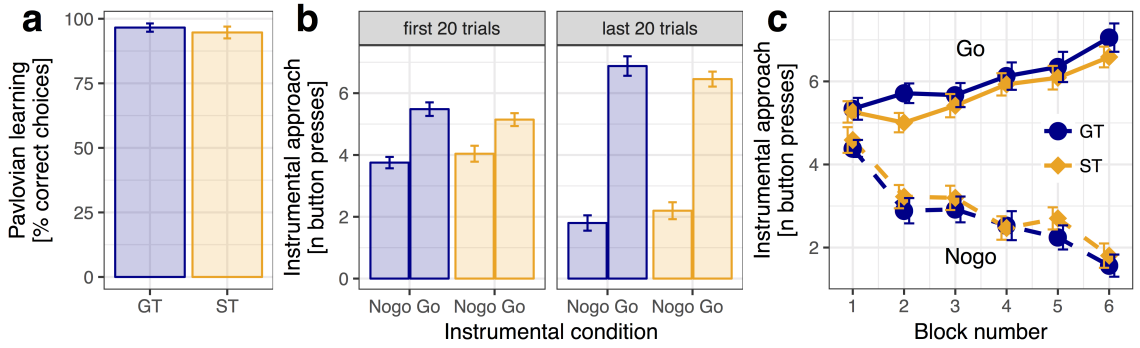
In sign-trackers, the model assuming the pupil is dilated by high model-free value explained the data better ($BIC = 7457.1$) than the model assuming model-based uncertainty dilates the pupil ($BIC = 7460.1$), suggesting that pupil dilation in sign-trackers is influenced by Pavlovian value predictions [11]. In goal-trackers, to the contrary, the model assuming pupil dilation relates to model-based uncertainty ($BIC = 8023.6$) outperformed the model-free CS value model ($BIC = 8032.4$), suggesting pupil dilation in goal-trackers is driven by aspects of model-based learning [12]. Comparison of model predictions with observed pupil sizes (see Fig. 2e+f) shows the model assuming value-based pupil dilation captures the continuous increase in the CS value effect on pupil size across trials in sign-trackers, whereas the model assuming uncertainty-based pupil dilation captures the decrease in pupil size across trials in goal-trackers.

Learning rate estimates in the value model yielded $\alpha = 0.056$ (logit transform: $a = -2.83$), though there was large posterior uncertainty (95% posterior credible interval for the learning rate scale: $\alpha = [0.0002\ 0.94]$; on the logit-transformed scale: $\sigma_a^{post} = 2.76$). Estimates of the learning rate for model-based state learning were better constrained $\eta = 0.20$ ($n = -1.40$; posterior standard deviation: $\sigma_n = 0.82$). While we note that the data was noisy, did not support a random effects analysis, and as such these estimates should be taken with caution, we also note a similarity between these estimates and those emerging from the fMRI and gaze analyses.

2.4 Behavioral results

2.4.1 Forced choice between Pavlovian CSs

Performance in forced choices between Pavlovian CSs based on their value (see Supplementary Fig. 4a) was close to ceiling for GT (97.8% correct answers, $SD_{subjects} = 9.2, SE = 1.5$, better than chance: $p_{bootstrap} < .001, 99.9\%\ CI = [86\ 99.6]$) and for ST ($95.2\%$ correct, $SD_{subjects} = 14.0, SE = 2.2$, better than chance: $p_{bootstrap} < .001, 99.9\%\ CI = [80\ 98.8]$), and did not significantly differ



Supplementary Figure 4 | **Pavlovian and instrumental learning in sign-trackers (ST, orange) and goal-trackers (GT, blue). a.** Percent correct forced choices between Pavlovian CSs assesses Pavlovian learning success. **b+c.** Performance during instrumental non-/approach training as assessed via the number of button presses. **b.** Instrumental performance during the first 20 trials (left) versus the last 20 trials (right) of instrumental training is displayed for instrumental conditions nogo (left bars) and go (right bars) for goal-trackers (left bars) and sign-trackers (right bars). **c.** Instrumental performance across six blocks, each containing one sixth of all individual training trials, is displayed for instrumental conditions go (solid lines) and nogo (dashed lines). **a-c.** Error bars are SEM.

between ST versus GT ($p_{bootstrap} > .1, b = 2.6\%, SE = 2.6, 95\%\ CI = [-1.8\ 8.6]$), suggesting that both groups successfully learned Pavlovian values.

2.4.2 Instrumental conditioning

ST and GT needed a similar number of trials to achieve the instrumental learning criterion (ST vs. GT: Mean [SD] = 75.8 [21.4] vs. 77.9 [23.0] trials; group difference: $p_{bootstrap} > .1, b = 2.1, SE = 4.8, 95\%\ CI = [-7.2\ 11.3]$). In both groups, approach and non-approach conditions differed in the number of button presses already in the first twenty trials (Supplementary Fig. 4b; ST: $t_{42} = 3.49, p = .001, b = 0.99, SD_{subjects} = 1.87, SE = 0.28, 99\%\ CI = [0.23\ 1.76]$; GT: $t_{42} = 6.46, p < .001, b = 1.83, SD_{subjects} = 1.85, SE = 0.28, 99.9\%\ CI = [0.83\ 2.83]$). This effect was stronger in GT compared to ST ($t_{84} = 2.08, p = .041, b = 0.83, SE = 0.40, 95\%\ CI = [0.03\ 1.63]$). In the last 20 trials the go/nogo effect was strong in GT ($p_{bootstrap} < .001, b = 5.14, SD_{subjects} = 2.88, SE = 0.44, 99.9\%\ CI = [3.43\ 6.35]$) and in ST ($t_{42} = 14.16, p < .001, b = 4.49, SD_{subjects} = 2.08, SE = 0.32, 99.9\%\ CI = [3.37\ 5.61]$), with no significant difference between ST and GT ($p_{bootstrap} > .1, b = 0.64, SE = 0.54, 95\%\ CI = [-0.47\ 1.63]$), indicating similar learning in both groups.

2.4.3 PIT

We found that during PIT instrumental response rate increased with appetitive and decreased with aversive Pavlovian value (see Fig. 3a,b; [13, 14]), reflecting a significant PIT effect in ST ($p_{bootstrap} < .001, b = 0.83, SD_{subjects} = 1.38, SE = 0.21, 99.9\%\ CI = [0.30\ 1.72]$), which was stronger ($p_{bootstrap} < .05, b = 0.49, SE = 0.26, 95\%\ CI = [0.09\ Inf]$) than in GT ($p_{bootstrap} < .001, b = 0.34, SD_{subjects} = 0.89, SE = 0.14, 99.9\%\ CI = [0.09\ 1.2]$). The PIT effect was in-

dividually significant in a higher percentage of ST (25.6%) compared to GT (9.8%; difference: $p_{bootstrap} < .05, b = 15.8, SE = 8.3, 95\%\ CI = [1.6\ Inf]$; see Fig. 3a, inset).

Moreover, we examined whether instrumental training affected instrumental performance during the PIT phase. Individuals who required less training trials had higher correct responses during the PIT phase ($t = -3.9, p < .001, b = -0.003, SE = 0.0008, 95\%\ CI = [-0.005\ -0.001]$). However, this did not affect the strength of the PIT effect per subject, i.e., it did not interact with the differential effect of Pavlovian value ($p = .16$). In addition, the difference in PIT between sign- and goal-trackers remained significant ($p < .05$) when we controlled for the number of training trials. Hence, although instrumental training does (as expected; c.f. [13]) affect instrumental performance during PIT, it does not relate to the actual PIT effect itself. This is in keeping with the notion that Pavlovian influences reflect contribution of a Pavlovian control system separate from the instrumental one.

Next, we tested whether subjects sometimes responded in a deterministic manner, which may indicate explicit response strategies rather than implicit effects of Pavlovian cues. As a first approach to assess deterministic responses, we considered subjects who performed all responses correctly, i.e., who collected all go-shells and left behind all no-go shells. There were a total 9 such perfect subjects: 6 out of 41 GT and 3 out of 43 ST, reflecting no significant group-difference ($p = .43$). In an additional analysis, we removed these 9 perfect subjects from the data and repeated the analysis of PIT effects. The results showed that the group-difference between sign- and goal-trackers in the PIT effect was still significant ($p < .05$) as before.

However, the behaviour of these participants may still not be fully deterministic, as participants could collect or leave shells using different number of button presses. We therefore considered an alternative definition of deterministic behaviour: We looked at responses to each instrumental shell per subject separately, and tested whether all responses to this shell in a given subject involved exactly the same number of button presses (e.g., all zero). This indicates a stricter definition of deterministic responses. Across all subjects this occurred in 79 instances, 45 times in GT and 34 times in ST, reflecting no significant group difference ($p > .1$). Next, we removed all instances with such deterministic behaviour from the data and re-computed the PIT effects. In this analysis, the PIT effect was still significantly stronger in sign- than in goal-trackers as before ($p < .05$), supporting a more implicit effect of Pavlovian value on instrumental behaviour.

2.5 fMRI results

2.5.1 Difference in RPE response between sign- and goal-trackers

There was a significant difference in the RPE response between sign-trackers and goal-trackers, $F(1, 75) = 10.88, SVCp_{FWE} = .026, [12\ 6\ -14], \eta_p^2 = .122, 95\%\ CI = [0.020\ 0.266]$ (Fig. 4a+b). We also studied the RPE separately for wins versus losses, and found a significant appetitive RPE signal in sign-trackers (average signal in NAc VOI: $t_{38} = 2.15, p = .019, b = 0.087, SE = 0.040, 95\%\ CI = [0.005\ 0.169]$), but not in goal-trackers ($t_{38} = -0.04, p = .516$; difference: $t_{75.9} = 1.53, p = .065, b = 0.089, SE = 0.058, 95\%\ CI = [0.008\ \infty]$; also see Supplementary Fig. 5). The aversive RPE response, however, was neither significant in goal-trackers nor in sign-trackers ($p > .1$).

One potential confound in the present analyses of group differences are visual effects. This could bias results because the group definitions were based on differences in fixation locations. We consider it unlikely that our highly specific computational measures of prediction errors are confounded by simple visual effects. Moreover, we test effects in brain regions known to encode

Supplementary Table 1 | Appetitive RPE signals for sign- and goal-trackers

VOI	Sign-trackers > 0			Goal-trackers > 0			ST $>$ GT			
	t(38)	*p*	lower CI	*t(38)*	*p*	lower CI	*df*	*t*	*p*	lower CI
NAc	2.15	0.019	0.019	-0.04	0.516	-0.072	75.9	-1.53	0.065	-0.008
VTA	1.90	0.033	0.006	-0.94	0.825	-0.122	63.2	-1.80	0.038	0.007
vmPFC	2.49	0.009	0.041	-0.10	0.541	-0.087	75.8	-1.88	0.032	0.015
Putamen	1.94	0.030	0.008	0.05	0.479	-0.059	72.8	-1.18	0.121	-0.023
Caudate	1.16	0.126	-0.016	-1.11	0.862	-0.113	71.5	-1.59	0.058	-0.004
Amygdala	2.86	0.003	0.052	-1.18	0.876	-0.146	74.3	-2.76	0.004	0.074

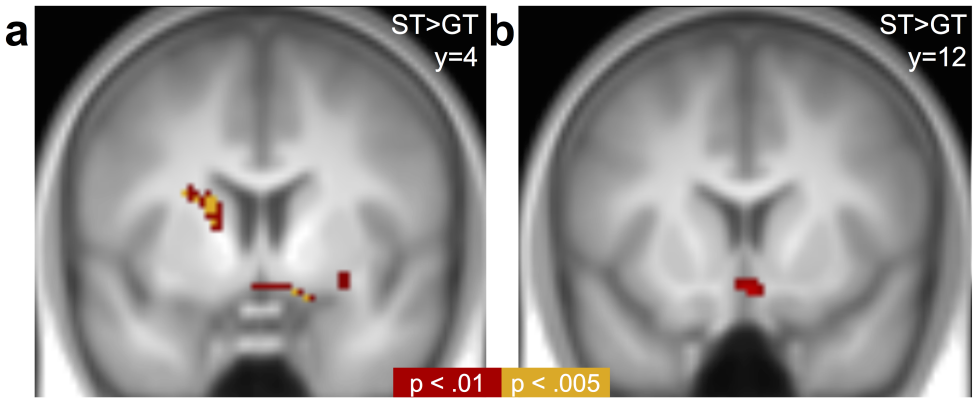
Note. Average appetitive RPE per VOI. "Lower CI" is the lower 95% confidence interval of a one-sided statistical test that the signal is larger than zero or that the signal is larger in sign- than in goal-trackers.

prediction errors rather than visual processing. However, we performed a control analysis to further test this potential confound. To this end, we coded two additional fMRI control regressors. The first coded as onsets each time a subject fixated on the CS stimulus, with the duration of the onset regressor given by the amount of time the subject was fixating on the CS. We expected that this regressor should capture visual processing of the CS. Indeed, maximal activation of this regressor was present in occipital regions. The second control regressor coded the same for fixations on the US. Importantly, both regressors did not show any group-difference in our a priori NAc VOI ($p_{uncorrected} > .05$). Next, we tested whether the group difference between sign- and goal-trackers in the reward prediction error signal in the a priori NAc VOI was still significant after controlling for the two visual regressors. Indeed, sign-trackers still showed a stronger RPE signal in the NAc than goal-trackers ($t(75) = 3.04, SVC\ p_{FWE} = .026$). This result supports our assumption that group-differences in RPE signals reflect differences in prediction error signaling rather than differences in visual perceptual processing.

2.5.2 Response in other brain regions

A key question was whether the difference in RPE signals between sign- and goal-trackers observed in the NAc was on average present also across other VOIs of the brain reward system. To test whether the difference in RPE signal was present on average, we extracted the average RPE signal for each of 5 additional VOIs from the brain reward system, including the ventral tegmental area (VTA), ventromedial prefrontal cortex (vmPFC), Putamen, Caudate, and Amygdala. Our hypothesis was that across these VOIs there should be a stronger RPE signal in sign-trackers than in goal-trackers, and we used an ANOVA with factors group (ST/GT) and VOI to test this. We found that across VOIs the RPE signal was stronger in sign-trackers than in goal-trackers ($F(1, 76) = 4.18, p = .044, \eta_p^2 = .01, 95\%\ CI = [0.00\ 0.04]$; see Supplementary Figure 6). However, the group-difference did not interact with VOI, suggesting the effect was stable across VOIs ($p = .363$). Moreover, across VOIs, the RPE signal was present in sign-trackers ($t_{76} = 1.893, p = .031, b = 0.035, SE = 0.019, 95\%\ CI = [0.004\ \infty]$) but not in goal-trackers ($p = .322$). These results support the stability of the RPE signal in sign-trackers and show that it is present across different regions of the brain reward system.

We further followed up on these results in exploratory analyses by studying appetitive RPE



Supplementary Figure 5 | **Appetitive RPE response in sign-trackers (ST) versus goal-trackers (GT). a+b.** Contrast showing stronger RPE signal in ST than in GT. Thresholds: $p < .01, k = 90$ (red) and $p < .005, k = 90$ (orange). Results are displayed for the peak group difference ($y = 4$; **a**) and for the peak signal in sign-trackers ($y = 12$; **b**).

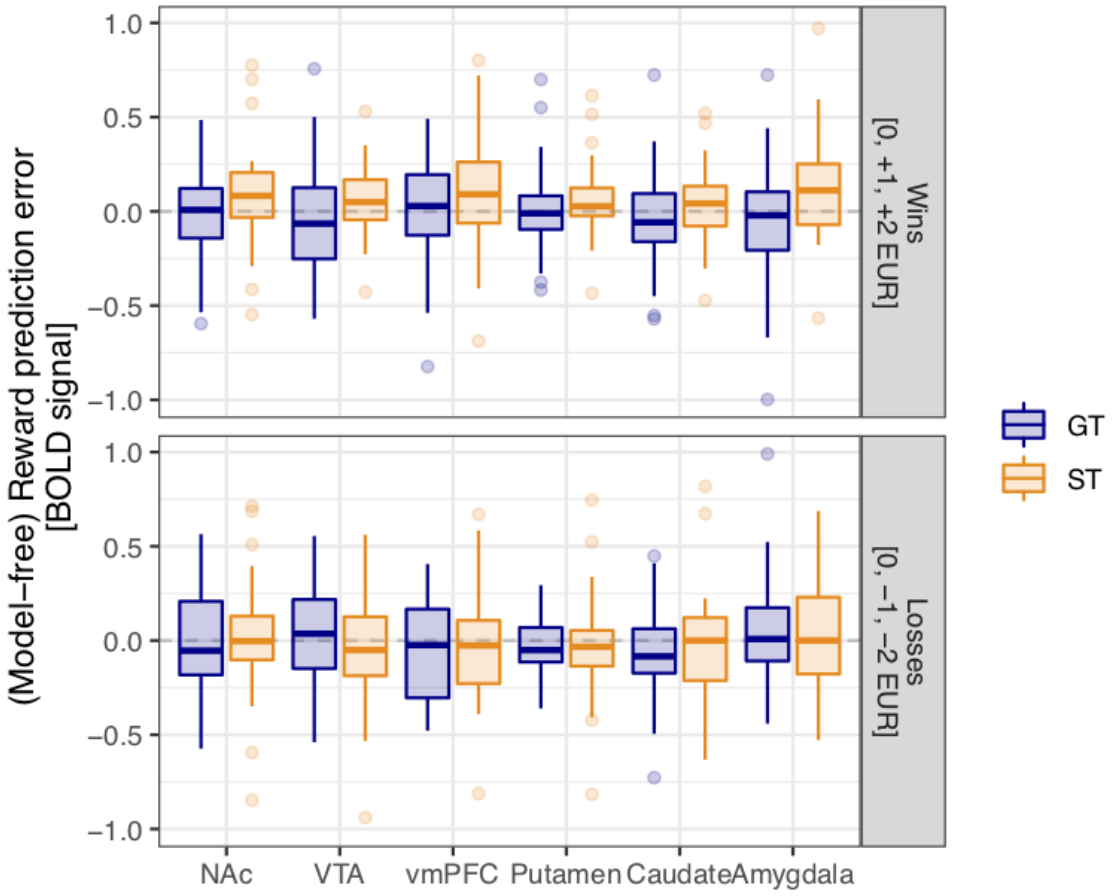
signals. First, we again used an ANOVA to test the average group-difference across VOIs, and again found a significant effect ($F(1, 76) = 5.40, p = .023, \eta_p^2 = 0.014, 95\%\ CI = [.00008\ 0.045]$), but no significant interaction ($p = .145$), indicating a stable group-difference across VOIs. Moreover, we followed up on this overall effect by exploring effects in individual brain regions. We found (Fig. 5) a wide-spread appetitive RPE signal in sign-trackers (all $p \leq .033$ except for caudate: $p = .126$), but not in goal-trackers (all $p \geq .479$), that was significantly stronger in sign- than in goal-trackers (all $p \leq .065$ except for putamen: $p = .121$). This result was observed across regions (see Supplementary Table 1). We also performed Holm correction for multiple comparisons, and found that Amygdala survived correction for the 6 tests: NAc: $p = .174$; VTA: $p = .160$; vmPFC: $p = .160$; Putamen: $p = .174$; Caudate: $p = .174$; Amygdala: $p = .024$. The aversive RPE signal, however, was neither present in sign-trackers, nor in goal-trackers, and there was no significant group difference in any of the VOIs ($p - values \geq .124$).

An analysis including data from all VOIs and from appetitive and aversive RPE simultaneously showed an appetitive RPE signal across VOIs in sign-trackers ($t_{145.8} = 2.54, p = .012, b = 0.071, SE = 0.028, 95\%\ CI = [0.016\ 0.126]$), but not in goal-trackers ($p = .471$), and not for aversive RPE (ST: $p = .591$; GT: $p = .436$). In sign-trackers, the appetitive RPE signal was stronger than the aversive RPE across VOIs ($t_{76} = 1.99, p = .051, b = 0.086, SE = 0.043, 95\%\ CI = [-0.0003\ 0.173]$). Moreover, sign-trackers showed a stronger appetitive RPE signal than goal-trackers across VOIs ($t_{146} = 2.31, p = .022, b = 0.091, SE = 0.040, 95\%\ CI = [0.013\ 0.169]$).

Last, we conducted exploratory whole-brain analyses, but found no evidence for group-differences in RPE signals.

2.5.3 Varying model-free learning rates

We next tested whether the RPE signal - assessed across all VOIs combined - was stable for different values of the free learning rate parameter. We first looked at the a priori analysis involving wins and losses, and performed an ANOVA involving factors group (ST/GT) and VOI (all 6 VOIs) for each of



Supplementary Figure 6 | **Distributions of average appetitive and aversive RPE responses in sign-trackers (ST) versus goal-trackers (GT) in different VOIs.**

a number of different values of the learning rate (0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6). We found that the difference in the RPE signal between sign- and goal-trackers was highly stable across learning rates: we found a significant group difference for the following values of the learning rate: 0.1: $p = .034$, 0.2: $p = .039$, 0.3: $p = .046$, 0.4: $p = .048$, 0.5: $p = .0495$, 0.6: $p = .04998$. For a learning rate of 0.001, the difference was not significant ($p = .175$). Moreover, the interaction between group (ST/GT) and VOI was not significant for any of the learning rates ($p > .4$), indicating the difference in RPE signals between groups was stable across the different tested VOIs. These results strongly support the reliability and generalizability of our main finding that model-free reward prediction error signals were stronger in sign-trackers than in goal-trackers.

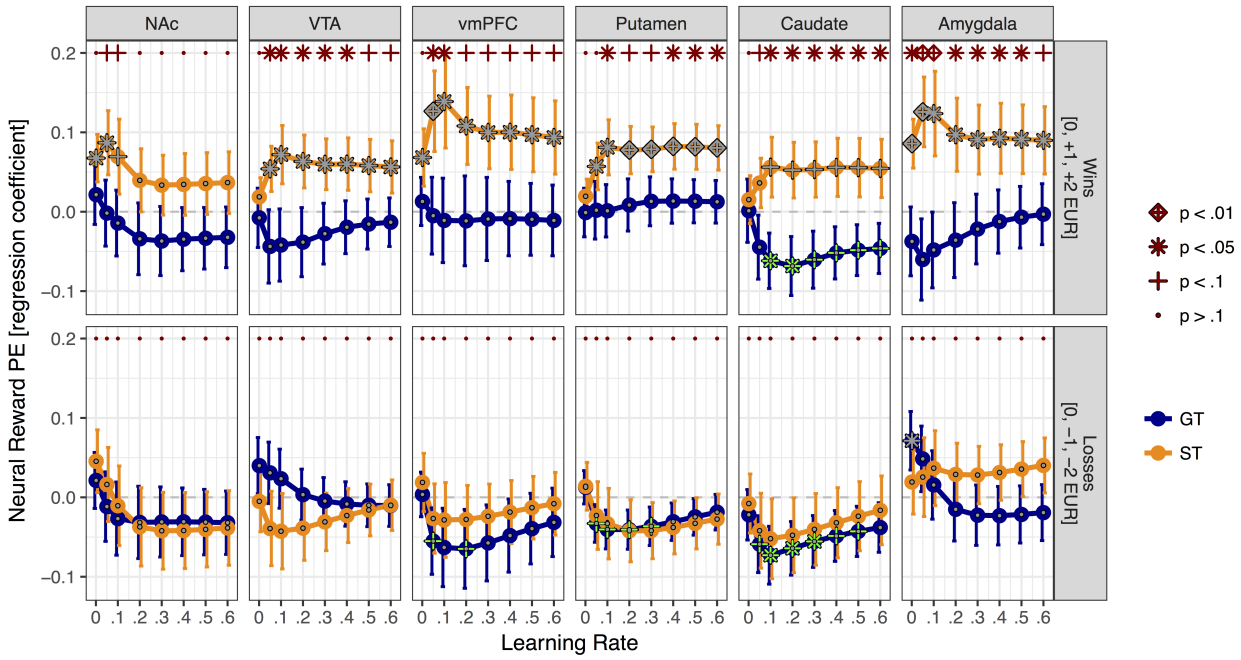
We also performed the same analysis for gains only, and again found that the group-difference in the RPE signal between sign- and goal-trackers was significant for most of the tested learning rates: 0.05 ($p = .029$), 0.1 ($p = .025$), 0.2 ($p = .039$), 0.3 ($p = .048$), and 0.4 ($p = .048$); and the effect was marginal ($p < .06$) for learning rates 0.5 ($p = .052$) and 0.6 ($p = .056$).

We next tested whether the observations about different learning rates to compute the RPE remained apparent when analyzing appetitive versus aversive RPE signals. To this end, we estimated the average appetitive versus aversive RPE signals per VOI for each of a set of different learning rates: $\alpha = [0.001,\ 0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6]$. The results, displayed in Supplementary Fig. 7, showed that the key result of a stronger appetitive RPE response in sign-trackers compared to goal-trackers (cf. [3]) was numerically stable for individual VOIs for at least some different values of the learning rate parameter. The appetitive RPE signal was most pronounced for small learning rates of ($\alpha = 0.05$ or $\alpha = 0.10$) in NAc, VTA, vmPFC and Amygdala, while for other regions (Putamen and Caudate) the effect was rather stable across different learning rates. The large BOLD effects at small learning rates are in keeping with the learning rate estimated from the pupil size ($\alpha_{pup} = 0.06$) [10]. However, for the smallest learning rate (e.g., $\alpha = 0.001$), the appetitive RPE signal was not significant in VTA, Putamen, and Caudate.

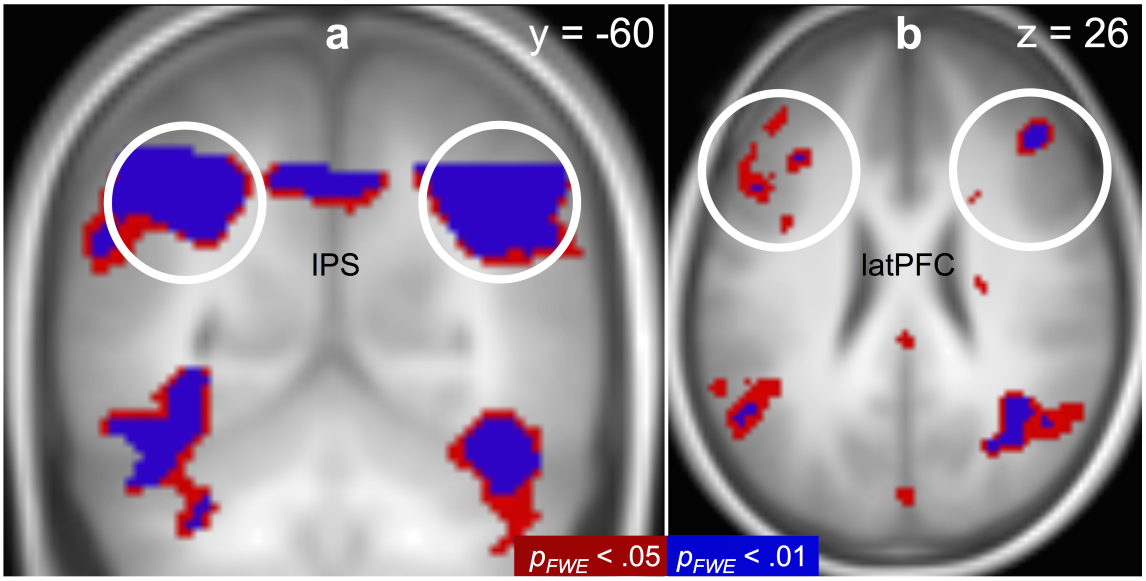
Of note, aversive RPE signals were neither present in sign- nor in goal-trackers with the exception of one single significant effect that did not survive correction for the 96 tests. Moreover, in goal-trackers we found some unexpected evidence for inverse RPE signals, mainly in Caudate. We currently do not understand this inverse signal, which may potentially be related to salience signals in goal-trackers.

2.5.4 State prediction error signals

State prediction errors (SPE) have been reported in intraparietal sulcus (IPS) and lateral prefrontal cortex (latPFC), and SPE were linked to model-based behavioural control specifically for the IPS [16]. Based on previous theorizing that goal-trackers rely more on model-based Pavlovian conditioning [5, 6] we here tested SPE signals in both groups. For sign- and goal-trackers combined, we found strong state prediction error signals in the *a-priori* VOIs of IPS (average response: $t(111) = 8.4, p < .0001, b = 1.20, SE = 0.14, 95\%\ CI = [0.92\ 1.48]$; peak-voxel: $t(75) = 8.6,\ SVC\ p_{FWE} < .001, [-30\ \ -60\ 40]$) and the lateral PFC (average response: $t(111) = 6.0, p < .0001, b = 0.858, SE = 0.143, 95\%\ CI = [0.575\ 1.14]$; peak-voxel: $t(75) = 6.0,\ SVC\ p_{FWE} < .001, [-50\ 16\ 40]$). Peak activations in both VOIs were also significant at a whole-brain corrected level ($p_{FWE} < .05$, see Supplementary Fig. 8). We found an interaction of group (ST/GT) with VOI ($F(1, 76) = 5.34, p = .024, \eta_p^2 = 0.033, 95\%\ CI = [0.0001\ 0.106]$). Following up on this interaction using post-hoc tests, we found that in the IPS as expected the SPE signal was significantly stronger in goal-trackers than in sign-trackers (average response: $t(67) = 2.12, p = .019, b = 0.564, SE =$



Supplementary Figure 7 | **Appetitive and aversive neural RPE signals for a range of learning rates** $\alpha = [0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6]$. We extracted the average RPE response per VOI and learning rate. Red marks (located at the top of each panel) show results from two-sample t-tests (see legend for p-values) testing our a priori hypothesis [3, 15] that the RPE signal is larger in sign- than in goal-trackers (one-tailed). Grey marks (located on displayed data points) indicate results from one-sample t-tests (see legend for p-values) testing whether the RPE signal for wins (upper panels) or losses (lower panels) is larger than zero (one-tailed) in each group of sign- or goal-trackers and for each learning rate and VOI separately. Moreover, we show results from exploratory one-tailed t-tests testing for inverse RPE signals, i.e., stronger BOLD response for negative TD PE (green marks; located on displayed data points).

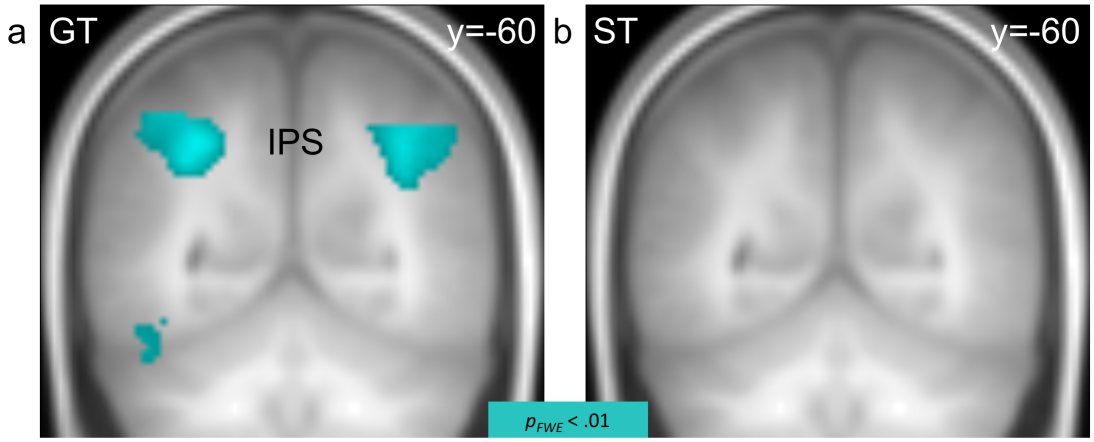


Supplementary Figure 8 | **State prediction error signals** in the intraparietal sulcus (IPS, **a**) and the lateral PFC (latPFC, **b**) for sign- and goal-trackers combined.

$0.266, 95\%\ CI = [0.121\ \infty]$; peak-voxel: $t(75) = 3.3, p_{unc.} < .001, [-46\ \ -36\ 44]$, see Fig. 6a+b and Supplementary Fig. 9), but the signal in the lateral PFC did not differ between groups ($p > .1$). These results support a stronger neural signature of model-based learning in goal-trackers, and thus provide support for computational models of sign- and goal-tracking [5, 6].

We note that the state prediction error regressor was perfectly orthogonal to the reward prediction error regressor in our experimental design ($r = 0$). We also tested whether reward prediction error signals would be observed in the IPS or lateral PFC in sign-trackers or in goal-trackers, and whether the groups differed in their reward prediction error signals in these VOIs. However, we found that neither sign- nor goal-trackers showed a reward prediction error signal in IPS ($p > .7$) nor in lateral PFC ($p > .7$) and that there was no evidence for a group-difference in either VOI ($p > .7$). These results underline the double dissociation of learning signals between sign- and goal-trackers.

Again, we performed additional analyses to control for visual perceptual effects when testing for the state prediction error signal in the IPS, using the same analysis approach described for the reward prediction error. We again found that maximal visual responses were located in occipital regions, and that sign- and goal-trackers did not differ in their visual responses in the IPS ($p_{FWE} > .2$). Crucially, the group difference in SPE signaling in the IPS was still significant after controlling for visual responses as goal-trackers still had stronger SPE signals in IPS than sign-trackers ($t(76) = 2.03, p = .046$), supporting our conclusion that state prediction error signals in IPS were stronger in goal-trackers than in sign-trackers, and that this was not confounded by group differences in visual perceptual processing.



Supplementary Figure 9 | **State prediction error signals** in the intraparietal sulcus (IPS) in goal-trackers (**a**) and in sign-trackers (**b**). $p_{FWE} = .01$, $k = 0$

2.6 Intermediate group between sign- and goal-trackers

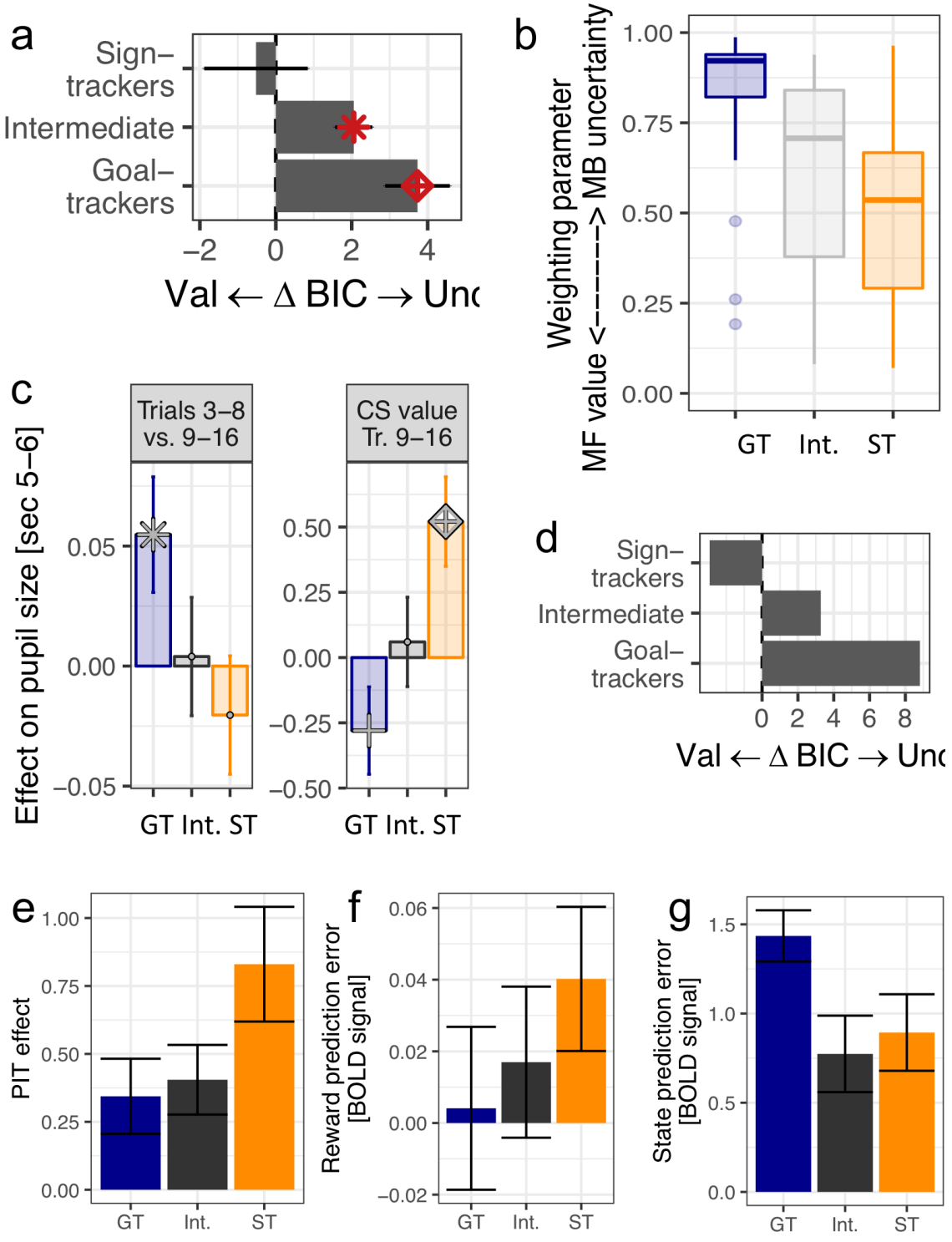
Throughout this work, we have compared the groups of sign-trackers and goal-trackers. However, in our group definition based on gaze there was also a third group of "intermediates", who showed near-zero effects of CS value on the gaze index (see Fig. 1b). We here tested whether this group of intermediates also showed intermediate results on a range of tested experimental measures. Indeed, as is displayed in Supplementary Figure 10, we found that intermediates had intermediate values in measures of computational models of gaze, in statistical and computational pupil analyses, in Pavlovian-instrumental transfer, in the NAc reward prediction error signal, and in the state prediction error signals in IPS.

2.7 Alternative definitions of sign- and goal-tracking

2.7.1 Distinguishing STs and GTs based on CS fixations

Our approach to defining sign- and goal-trackers was based on the CS value effect on the gaze index. We here implemented an additional analysis approach and defined sign- and goal-trackers not based on the linear CS value effect on the gaze index, but instead on the probability to fixate the CS, i.e., treating fixations on the background and on the US as the same. In this alternative definition, 36 out of the original 43 sign-trackers, and 37 out of the 43 original goal-trackers were again classified as sign-/goal-trackers, respectively, while the others switched to being defined as intermediates.

With this alternative definition of sign- and goal-trackers, we repeated our key analyses on Pavlovian-instrumental transfer (PIT), neural reward prediction error signals (RPE), and on state prediction errors (SPE) and again found the same group differences (see Supplementary Fig. 11a-c). Sign-trackers defined based on CS-related fixations again had a stronger PIT effect compared to goal-trackers ($p_{bootstrap} < .05, b = 0.35, SE = 0.23, 95\%\ CI = [0.02\ Inf]$). Moreover, CS-



Supplementary Figure 10 | **Analysis results for the intermediate group**. **a+b**, Computational modelling results for the gaze index, including relative BIC (**a**), and the mixing weight between model-free and model-based control (**b**). **c**, Pupil analyses including the effect of trials (left) and of CS value at the end of learning (right). **d**, Computational modelling results for pupil dilation, showing relative BIC. **e**, Average PIT effect for each group. **f**, Average reward prediction error response (BOLD signal) in the NAc per group. **g**, Average state prediction error response (BOLD signal) in the IPS per group.

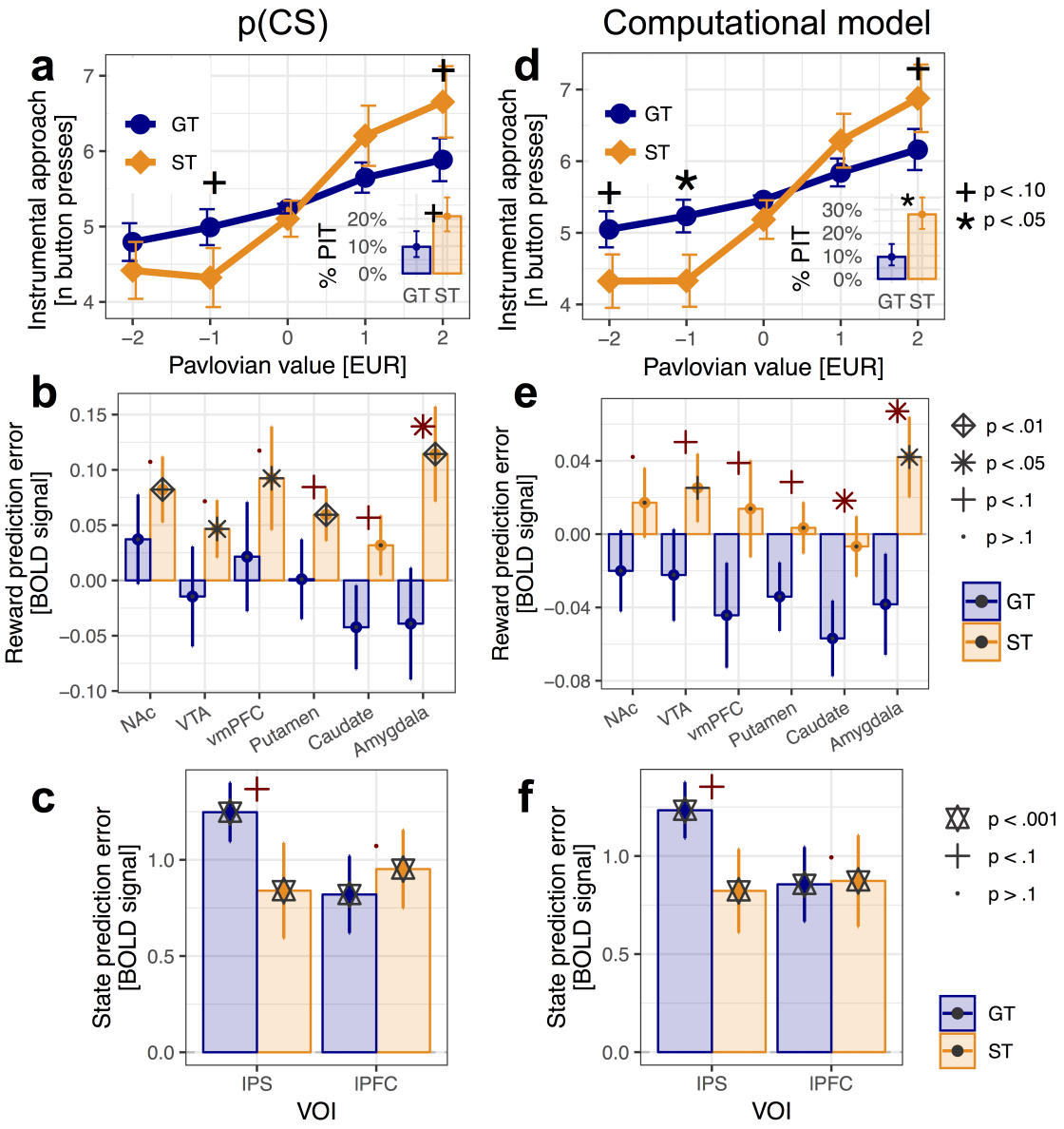
defined sign-trackers had stronger RPE signals in the NAc compared to goal-trackers (voxel-wise difference: $F(1,75) = 14.11, SVCp_{FWE} = .007, [12\ 6\ -14], \eta_p^2 = 0.152, 95\%\ CI = [0.036\ 0.301]$). Sign-trackers also showed an appetitive RPE signal in NAc (average signal: $t_{38} = 2.76, p = .004$), and in the other VOIs ($p - values < .05$, Caudate: $p = .123$), which was absent in goal-trackers ($p - values > .1$). In an analysis testing all VOIs simultaneously, sign-trackers showed an appetitive RPE signal ($t_{76} = 2.33, p = .011, b = 0.07, SE = 0.03, 95\%\ CI = [0.01\ 0.13]$), which was stronger compared to goal-trackers ($t_{76} = 1.79, p = .039, b = 0.08, SE = 0.04, 95\%\ CI = [0.0003\ 0.16]$), who themselves showed no evidence for an appetitive RPE signal ($p = .844$). Last, we found that goal-trackers defined based on CS-related fixations had marginally stronger SPE signals compared to sign-trackers in IPS ($t_{63} = 1.40, p = .083, b = 0.408, SE = 0.291, 95\%\ CI = [0.078\ \infty]$), but not in lPFC ($p = .678$, interaction VOI x group: $F(1,76) = 4.63, p = .03, \eta_p^2 = 0.029, 95\%\ CI = [0.000\ 0.099]$).

2.7.2 Distinguishing STs and GS based on computational model fits

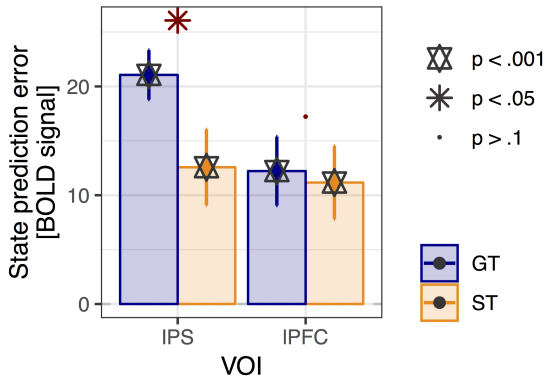
We next fitted a computational model to the gaze data assuming dual control by uncertainty and by a model-free Pavlovian conditioned response bias from CS value (section "Using computational modeling to define sign- and goal-trackers"). We assumed that the weight of CS value and specifically its size and direction defines sign- versus goal-trackers: we defined sign- vs. goal-trackers as the tertiles of subjects with the most positive vs. most negative parameter estimate for the weight of CS value on gaze direction (i.e., b_V^{gaze}). The estimated weight parameter was highly correlated with the linear CS value effect on gaze direction that we had used in the original definition of sign- and goal-trackers (across subjects: Spearman's $\rho = .84, S = 56388, p < .001$). Moreover, a substantial number of participants were assigned to the same groups: 32 out of the 43 sign-trackers, 35 out of the 43 goal-trackers were again classified as ST/GT, while the remaining subjects became intermediates, and only one goal-tracking subject switched to being defined as sign-tracker in the computational analysis.

We tested whether the key behavioural and imaging results were stable when using this computational definition of sign- and goal-trackers. The main results from these analyses are displayed in Supplementary Figure 11d-f. First, we tested the hypothesis that the weight of uncertainty on gaze direction, i.e., β_U^{gaze}, was larger in computationally-defined goal-trackers ($M = 0.76, SE = 0.09$) than in computationally-defined sign-trackers ($M = 0.51, SE = 0.09$), and found this to be the case (significant difference: $t_{84} = 1.76, p = .041, b = 0.24, SE = 0.13, 95\%\ CI = [0.01\ \infty]$), supporting a stronger reliance on uncertainty-based attention in goal-trackers.

Second, we found the PIT effect was stronger in the computationally defined STs than GTs ($p_{bootstrap} < .05, b = 0.42, SE = 0.22, 95\%\ CI = [0.10\ \infty]$) and more frequently individually present ($z = 2.08, p = .038, b = 1.30, SE = 0.63, 95\%\ CI = [0.14\ 2.65]$; see Supplementary Fig. 11d). Third, neural signatures of value and state learning replicated. The neural RPE signal (computed across wins and losses) was stronger in computationally-defined sign- than goal-trackers in the NAc (voxel-wise difference: $F(1,74) = 9.38, SVCp_{FWE} = .003, [14\ 4\ -14], \eta_p^2 = 0.107, 95\%\ CI = [0.013\ 0.248]$), and across all tested VOIs (mean signal: $t_{75} = -2.11, p = .019, b = -0.05, SE = 0.02, 95\%\ CI = [-0.09\ -0.01]$). Third, there were again strong SPE signals in IPS and lPFC for both groups ($p - values < .001$), with a (now marginal) interaction between group (sign-/goal-trackers) x VOI ($F(1,75) = 3.01, p = .09, \eta_p^2 = 0.019, 95\%\ CI = [0.000\ 0.082]$) indicating that the SPE signal was stronger in sign-trackers than in goal-trackers in IPS ($t_{66} = 1.60, p = .057, b = 0.41, SE = 0.26, 95\%\ CI = [-0.02\ \infty]$), but not in lateral PFC ($p = .523$). The results replicate our



Supplementary Figure 11 | **Alternative definitions of sign- and goal-tracking. a-c.** Sign- and goal-trackers are defined based on the linear CS value effect on the probability to fixate the CS. **d-f.** Sign- and goal-trackers are defined based on a computational model of gaze direction. **a + d.** Number of instrumental button presses during the PIT task as a function of Pavlovian CS background value for alternatively defined sign- versus goal-trackers. **a + d insets.** Percentage of subjects with an individually significant PIT effect. **b + e.** Reward prediction error BOLD signal in different VOIs. **c + f.** State prediction error BOLD signal in different VOIs. **a-f.** Error bars are SEM.



Supplementary Figure 12 | **State prediction errors from Bayesian model-based learning.**

main findings on differences between sign- and goal-trackers using a computational definition of their gaze control.

2.8 Bayesian model of model-based learning and uncertainty

We additionally used a Bayesian model of model-based learning about state transitions. This Bayesian model allowed to directly compute trial-by-trial uncertainty, and to use this to predict gaze (cf. equation 3). The model has a parameter for a priori observations about transitions, which we estimated from the gaze index data. Based on this model, we again found that goal-trackers relied more strongly on the model-based system than on the model-free system ($t_{84} = 3.77, p = .0003, b = 4.49, SE = 1.19, 95\%\ CI = [2.12\ 6.85]$). Likewise, evidence in sign-trackers was again significantly shifted towards the model-free Pavlovian response bias (interaction group x model: $F(1, 84) = 5.52, p = .021, \eta_p^2 = 0.031, 95\%\ CI = [0.0004\ 0.098]$), and showed similar support for model-free and model-based control ($p = .660, b = -0.53, SE = 1.19$). Next, we also looked at state prediction errors computed in the Bayesian model-based system, and how they are coded in IPS and in lPFC. We again found (see Supplementary Fig. 12) a significant interaction of VOI x group ($F(1, 76) = 4.43, p = .039, \eta_p^2 = 0.027, 95\%\ CI = [0.000\ 0.097]$): goal-trackers showed a stronger SPE signal than sign-trackers in IPS ($t_{65} = 2.04, p = .023, b = 8.49, SE = 4.17, 95\%\ CI = [0.17\ 16.81]$) but not in lPFC ($p = .410$).

2.9 Learning from wins versus losses

While studies on sign- and goal-trackers usually focus on appetitive conditioning (e.g., [2, 3]), CSs predicting aversive USs are also known to elicit sign-tracking responses [17]. Based on theoretical considerations that learning should manifest in differences between appetitive and aversive predictions we had designed the task to test differences in learning between wins and losses. Accordingly, the design was not optimized to test differences to neutral trials. For instance, we did not equalize the number of neutral and reward trials. However, the fact that the fMRI analyses showed specific coding only of appetitive RPEs (see section 2.5) prompted us to test whether effective learning as measured in gaze direction and pupil dilation also differed between wins and losses. We investi-

gated this question using computational modeling. We compared computational models assuming control by CS value from a single learning rule and a single weight of CS value with alternative models assuming a difference between wins and losses in (i) the learning rate or (ii) the weight of CS value on the dependent variable. Neither gaze nor pupil dilation data suggest a distinction in the learning from wins and losses:

2.9.1 Gaze index

We first estimated (regression) weights of gaze direction on CS value for wins and losses separately, i.e., allowing two separate weights instead of one, but keeping the same learning rule across wins versus losses. The analysis showed that allowing different weights for wins versus losses (i.e., $b^{\text{gaze}}_{V,wins}$ and $b^{\text{gaze}}_{V,Loss}$) did not improve the overall fit of the model compared to the baseline model, which included only a single identical weight parameter for wins and losses (i.e., b^{gaze}_{V}). This was the case across all subjects (baseline model value: $Mean$ Log Likelihood (LL)$= -42.5,\ Mean$ BIC $= 112.7$; dual weights model: $Mean\ LL = -41.2,\ Mean$ BIC $= 119.2$; $t_{128} = 20.06, p < .001, \Delta LL = 1.3,\ \Delta$BIC $= 6.6, SD_{subjects} = 3.7, SE = 0.3, 99.9\%\ CI = [5.5\ 7.7]$), and also for the group of sign-trackers, for whom the value model prevailed (baseline value model: $Mean\ LL = -50.6,\ Mean$ BIC $= 125.3$; dual-weight model: $Mean\ LL = -49.6,\ Mean$ BIC $= 131.5$; $t_{42} = 18.26, p < .001, \Delta LL = 1.0,\ \Delta$BIC $= 6.1, SD_{subjects} = 2.2, SE = 0.3, 99.9\%\ CI = [4.9\ 7.3]$).

Second, we tested a model where the (regression) weight b^{gaze}_{V} was identical across wins and losses, but where different learning rates where assumed for the wins versus losses (i.e., $\alpha^{\text{gaze}}_{Win}$ and $\alpha^{\text{gaze}}_{Loss}$). We found that the baseline model explained the data better compared to this dual learning rate model, both for all subjects (baseline value model: $Mean\ LL = -42.5,\ Mean$ BIC $= 112.7$; dual learning rates model: $Mean\ LL = -42.3,\ Mean$ BIC $= 121.3$; $t_{128} = 95.16, p < .001, \Delta LL = 0.3,\ \Delta$BIC $= 8.7, SD_{subjects} = 1.0, SE = 0.1, 99.9\%\ CI = [8.4\ 9.0]$) and for the group of sign-trackers specifically (baseline value model: $Mean\ LL = -50.6,\ Mean$ BIC $= 125.3$; dual learning rates model: $Mean\ LL = -50.2,\ Mean$ BIC $= 132.7$; $t_{42} = 36.02, p < .001, \Delta LL = 0.4,\ \Delta$BIC $= 7.4, SD_{subjects} = 1.3, SE = 0.2, 99.9\%\ CI = [6.7\ 8.1]$). These results suggest a single learning rate and a single weight of CS value across wins and losses best explains the patterns of gaze.

2.9.2 Pupil dilation

We moreover used the same computational modeling approach conducted for gaze direction to analyze models of pupil dilation during Pavlovian conditioning. Again, we used our original reinforcement learning value model and fitted two alternative models to the pupil size data. The first model estimated weights of pupil size on CS value for wins and losses separately, while assuming the same learning rate for wins and losses. Allowing the value weights to differ between wins versus losses ($b^{\text{pupil}}_{V,Wins}$ and $b^{\text{pupil}}_{V,Loss}$) did not explain the pupil dilation better than the baseline model assuming a single weight (b^{pupil}_{V}) across all subjects (baseline model value: $LL = -11516.88,\ BIC = 23061$; dual weights model: $LL = -11513.75,\ BIC = 23064;\ \Delta LL = 3.13,\ \Delta\ BIC = 3$), and also for sign-trackers, where the value model was the winning model (baseline value model: $LL = -3716.69,\ BIC = 7457$; dual-weight model: $LL = -3716.69,\ BIC = 7465;\ \Delta LL = 0.00,\ \Delta\ BIC = 8$).

We tested a second model with identical weight parameter b^{pupil}_{V}, but with separate learning rates for the wins versus losses ($\alpha^{\text{pupil}}_{Win}$ and $\alpha^{\text{pupil}}_{Loss}$). The baseline model received more support

from the data than this dual learning rate model, when studying all subjects (baseline value model: $LL = -11516.88,\ BIC = 23061$; dual learning rates model: $LL = -11516.84,\ BIC = 23070;\ \Delta LL = 0.04,\ \Delta\ BIC = 9$) and when looking at sign-trackers only (baseline value model: $LL = -3716.69,\ BIC = 7457$; dual learning rates model: $LL = -3716.65,\ BIC = 7465;\ \Delta LL = 0.04,\ \Delta\ BIC = 8$).

References

1. Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. Differential encoding of losses and gains in the human striatum. *J. Neurosci.* **27,** 4826–31 (2007).
2. Garofalo, S. & di Pellegrino, G. Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. *Front. Behav. Neurosci.* **9,** 163 (2015).
3. Flagel, S. B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469,** 53–7 (2011).
4. Gottlieb, J. Attention, learning, and the value of information. *Neuron* **76,** 281–95 (2012).
5. 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).
6. 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).
7. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat. Neurosci.* **3,** 1218–23 (2000).
8. Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R. & Gottlieb, J. Reward modulates attention independently of action value in posterior parietal cortex. *J. Neurosci.* **29,** 11182–91 (2009).
9. Hogarth, L., Dickinson, A. & Duka, T. Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal and uncertainty. *Attention and associative learning: From brain to behaviour,* 71–97 (2010).
10. Wilson, R. C. & Niv, Y. Is model fitting necessary for model-based fMRI? *PLoS Comput. Biol.* **11,** e1004237 (2015).
11. Manohar, S. G. & Husain, M. Reduced pupillary reward sensitivity in Parkinson's disease. *NPJ Parkinson's disease* **1,** 15026 (2015).
12. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15,** 1040 (2012).
13. 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).
14. Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M. & Cools, R. Aversive Pavlovian control of instrumental behavior in humans. *J. Cogn. Neurosci.* **25,** 1428–41 (2013).
15. 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. *Neurosci.* **196,** 80–96 (2011).
16. Gläscher, J., Daw, N. D., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66,** 585–95 (2010).
17. Leclerc, R. Sign-tracking in aversive conditioning. *Learn. Motiv.* **11,** 302–17 (1980).