Reduced Pavlovian value updating alters decision-making in sign-trackers

Reduced Pavlovian value updating alters decision-making in sign-trackers

Luigi A.E. Degni*[, 1, 2], Lorenzo Mattioni*[, 1], Claudio Danti[1], Valentina Bernardi[1], Gianluca Finotti[3], Marco Badioli[1], Francesca Starita[1], Alireza Soltani[4], Giuseppe di Pellegrino[1], Sara Garofalo[1]

[1]Center for Studies and Research in Cognitive Neuroscience, Department of Psychology, University of Bologna, Cesena, Italy.
[2]International School of Advanced Studies, University of Camerino, Camerino, Italy.
[3]School of Psychological Sciences, University of London, Birkbeck.
[4]Department of Psychological and Brain Sciences, Dartmouth College, Hanover 03755, New Hampshire

*These authors contributed equally to this work.

Corresponding author: Giuseppe di Pellegrino; email: g.dipellegrino@unibo.it

Manuscript information: 5 figures, 3 tables, 35 pages

Keywords: Pavlovian-to-Instrumental Transfer; Sign-trackers; Reinforcement learning; Computational modelling; eye-tracking

Abstract

Successful reward-guided behavior relies not only on learning actions to obtain rewards but also on learning cues that predict the reward, which motivate and prepare the animal to pursue and consume it. Although these two types of learning-instrumental learning and Pavlovian conditioning-have been extensively studied, it remains unclear how the brain updates and arbitrates between these systems, especially when Pavlovian signals are irrelevant to decision making. To address this, we used eye-tracking, pupillometry, and computational modeling in a Pavlovian-to-Instrumental Transfer task with 60 humans (30 females), consisting of three phases: the Pavlovian phase (learning conditioned stimulus-outcome associations), the instrumental phase (learning response-outcome associations), and the transfer phase (testing Pavlovian bias on instrumental responses). Using this approach, we aimed to identify different types of learners and their strategies, especially how individual differences in sign-trackers versus goal-trackers influence Pavlovian bias. To that end, we used eye gaze data to categorize participants as sign- or goal-trackers, and found that although both groups learned the task, sign-trackers' performance was lower when exposed to Pavlovian cues, as they favored options based on their cue-outcome associations. Fitting data with multiple computational models revealed that participants dynamically arbitrated between values estimated through Pavlovian and instrumental systems. Importantly, lower performance in sign-trackers was due to slower updating of Pavlovian cue values during the transfer phase, not overweighting of Pavlovian cue values relative to instrumental action values. Overall, our study offers a computational framework for understanding inflexible decision making and potential interventions for disorders marked by maladaptive cue reactivity.

Significance Statement

Everyday decisions are shaped by the feedbacks received after making choices (instrumental learning) or predicted by the presence of cues —like images or sounds - around us (Pavlovian learning). Sometimes, these cues can lead us to make poor choices. Our study examines how these two learning systems interact, and why some individuals are more influenced by such cues than others. Using eye-tracking, pupillometry, and computational modeling, we show that people more drawn to reward-predicting cues update their beliefs more slowly, leading to biased decisions. These findings may offer new insight into why some individuals struggle with inflexible or maladaptive behavior, as seen in conditions like addiction and compulsive disorders.

Introduction

Value-based decision-making relies on the ability to leverage environmental cues to predict valuable outcomes and guide actions. Pavlovian learning mechanisms are fundamental to this process, shaping both attentional and goal-directed behavior (Soltani & Koechlin, 2022). These mechanisms are studied using Pavlovian autoshaping procedures (Brown & Jenkins, 1968), where a conditioned stimulus (CS) predicts a reward delivered in a different location (Berridge, 2000). Based on conditioned responses, two distinct learning profiles have been identified in non-human animals: sign-trackers, who approach the Pavlovian cue ("sign") upon its presentation, and goal-trackers, who move directly to the location of the reward ("goal") (Colaizzi et al., 2020). Notably, sign-trackers persist in approaching the cue even when these actions delay (Breland & Breland, 1961) or prevent reward retrieval (Hearst & Jenkins, 1974), suggesting a maladaptive prioritization of Pavlovian cues over instrumental goals.

Although the cue predicts the reward for both sign- and goal-trackers, it acquires a higher incentive value for the former (Robinson & Flagel, 2009). Rodent literature shows that conditioned responses in sign-trackers are more resistant to extinction (Ahrens et al., 2016) and that sign-tracking is linked to increased salience to reward cues but not to reward feedback (Duckworth et al., 2022). Accordingly, sign-trackers exhibit maladaptive behaviors linked to higher incentive value attribution, like increased impulsivity (Saunders & Robinson, 2010), attentional control deficits (Paolone et al., 2013), and propensity for addiction-like behaviors (Flagel et al., 2009). Translating these findings to humans, researchers have identified analogous learning styles by measuring oculomotor responses during Pavlovian tasks (for a review, see Colaizzi et al., 2020).

Despite parallels between human and animal models, two important questions remain. First, although animal models link sign-tracking to maladaptive cue-driven decision-making, a direct translation into suboptimal 'Pavlovian biases'—i.e., behavioral bias due to Pavlovian cues that reduce the capacity to maximize reward—in humans is lacking. Notably, previous evidence in humans (Garofalo & di Pellegrino, 2015; Schad et al., 2020) suggested enhanced cue-driven decision-making in sign-trackers. However, those studies assessed Pavlovian biases in which cues non-specifically invigorate responding, an effect known as general Pavlovian-to-Instrumental Transfer (PIT). By contrast, it is not yet established whether sign-tracking is associated with outcome-specific Pavlovian influences that shift choice among competing instrumental options, nor whether this bias holds if it results in suboptimal performance. Second, the computational mechanisms linking Pavlovian and instrumental learning remain unclear.

To address these questions, we employed a modified outcome-specific PIT paradigm featuring "poor" and "rich" choice options with a specific reward schedule (Badioli et al., 2024). Classic PIT tasks include three phases: Pavlovian learning, where participants learn CS-outcome associations; instrumental learning, where they learn response-outcome associations; transfer, where Pavlovian bias on instrumental responses is tested (Degni et al., 2025). Unlike classic PIT paradigms, participants here performed transfer with feedback, where reward delivery depended solely on instrumental actions (Pavlovian cues were task-irrelevant). Also, we implemented a baiting rule whereby the probability of being rewarded for each option increased the longer it was left unchosen. Under these conditions, reward matching (i.e., allocating responses in proportion to each option's nominal reward rate) was

the optimal strategy to maximize cumulative payoffs (Lau & Glimcher, 2005). Deviations from this pattern, such as responding equally across options or favoring the poor option, were therefore suboptimal as they led to a potential loss of reward. This foraging-like design allowed us to examine whether sign-trackers exhibit suboptimal Pavlovian bias.

Using computational modeling, we assessed two competing mechanisms through which motivational salience may lead to Pavlovian bias in sign-trackers (Ostlund & Marshall, 2021). First, the higher value attributed to reward-predictive cues could amplify their influence over instrumental decisions in sign-trackers. Alternatively, sign-trackers may fail to adequately update cue values in the transfer phase, leading to persistent suboptimal strategies due to incorrect estimates. To test these hypotheses, we fit trial-by-trial choice data using various computational models, followed by model selection to assess both Pavlovian and instrumental associative learning along with their dynamic interaction.

Materials and Methods

Participants

The sample size was established based on a power analysis conducted on MorePower 6.0 (Campbell & Thompson, 2012) for the planned 4x2 mixed-measures ANOVAs. Specifically, we used the following parameters: Repeated Measures design factors = 1 factor (CS) with 4 levels ($CS+_1$, $CS+_2$, CS-, No CS); Independent Measures design factors = 1 factor (Group) with 2 levels (sign-trackers, goal-trackers); effect size (η_p^2) = 0.08; and significance level = 0.05; power = 0.9. The effect size was estimated based on the minimum effect size (η_p^2) obtained by previous studies conducted with a Pavlovian-to-Instrumental Transfer task (Degni et al., 2022; Garofalo et al., 2019, 2020, 2021; Garofalo & di Pellegrino, 2015; Garofalo & Robbins, 2017;Degni et al., 2024). A-priori power analysis resulted in a minimum sample of 56. Here, the additional manipulation of the instrumental learning phase suggested a possible smaller effect size (Lakens, 2013). Hence, 60 volunteers with normal or corrected-to-normal vision and no history of neurological or psychiatric diseases were recruited. Data from 4 participants were unusable due to technical issues during eye-tracking recording. Two more participants did not provide usable eye-tracking data for classification as sign-trackers or goal-trackers because they did not reach the minimum fixation threshold (80% of the fixation time on the regions of interest; Cherkasova et al., 2024). Six additional participants were hence recruited to reach the planned sample size of 60 participants (30 females, mean age = 24.2; sd = 4.34 years; mean education = 16.4; sd = 1.9 years). The study was conducted in accordance with institutional guidelines and the 1964 Declaration of Helsinki and was approved by the Bioethics Committee of the University of Bologna (Prot. 201070, 24/09/2020).

Experimental Paradigm

We used a modified PIT paradigm (Cartoni et al., 2016), adapted from (Degni et al., 2025). The task involved three consecutive phases: Pavlovian learning phase, in which participants learned CS-outcome associations; Instrumental learning phase, in which participants

learned response-outcome associations; Transfer phase, in which Pavlovian bias on instrumental responses was tested.

An image of a slot-machine on a grey background was presented during all task phases, composed of two displays on the top and bottom and a fixation cross in the center (**Figure 1**). During the instrumental and transfer phases, two levers appeared, respectively on the right and left of the slot-machine, to represent the two available instrumental actions (**Figure 1b,c**). The task was programmed and run on OpenSesame v3.2 (Mathôt et al., 2012).

Pavlovian learning phase

In the Pavlovian learning phase, participants learned the associations between three CSs (i.e., fractal images balanced for luminance, complexity, and color saturation (Finke et al., 2021) and three corresponding outcomes. Two CSs were used as CSs+ ($CS+_1$ and $CS+_2$). In 80% of the trials, these CSs were respectively associated with two different rewarding outcomes (O_1 and O_2), which were highly and equally desirable food snacks tailored to each participant (Garofalo & di Pellegrino, 2015). In the remaining 20% of the trials, these CSs were associated with a non-rewarding outcome ("X"). The third CS served as CS-, associated with the non-rewarding outcome ("X") in all trials. The association between CSs and outcomes was counterbalanced across participants. On each trial (**Figure 1a**), an empty slot-machine was presented in the center of the screen (Intertrial Interval, ITI, 5000-6000 ms). Then, one of the three CSs appeared on its upper display for 5000 ms, followed by the image of the corresponding outcome (1000 ms) in the lower display, while the CS remained stable on the upper display. No levers were presented during this phase.

A series of blocks, each composed of 30 trials (i.e., 10 trials for each CS), were repeated until a learning criterion was achieved. Each block had an average duration of 6 minutes, followed by multiple-choice questions presented on the display to ensure that all CS-outcome associations were correctly established (i.e., “What food did you earn with this stimulus?”, repeated for each CS). The learning criterion consisted of correctly reporting the association between the three CSs and the three outcomes two times in a row. If the learning criterion was achieved, the participant moved on to the following phase; otherwise, the task was aborted after 4 wrong answers (i.e., from a minimum of 2 blocks to a maximum of 8).

Before this phase, a liking Likert scale ranging from 0 (not at all) to 9 (very much) was presented on the display to ensure comparable motivational value of the outcomes previously chosen by the participant, with two questions for each food snack: “How much do you usually enjoy eating it?” (i.e., general liking) and “How much would you like to eat it now?” (i.e., current wanting). Before and after this phase, the same liking 9-point Likert scale was presented for each of the three CSs, with the question “How much do you like this stimulus?” to ensure (a) that the CSs had comparable at the beginning of the experiment, and (b) that the acquisition of Pavlovian learning was effective at the end of the phase.

Instrumental learning phase

In the Instrumental learning phase, participants learned the association between two responses (R_1 or R_2) and two corresponding rewarding outcomes (O_1 and O_2). Crucially, the

task was structured such that participants had to learn an optimal decision-making strategy, i.e. the strategy that allowed maximizing the amount of reward earned.

On each trial (**Figure 1b**), an empty slot-machine was presented in the center of the display (ITI, 500-1500 ms). Then, two levers appeared respectively on the left and right side of the slot-machine, until the participant's response. The responses consisted of choosing one of the two levers presented on the display, that could be selected through two computer keys (i.e., "z" for the left and "m" for the right lever of the slot-machine). To avoid distraction and provide time pressure, participants were told to choose quickly after the appearance of the levers (which served as a "go" signal), otherwise, no outcome would appear, and the trial was aborted (Daw et al., 2011). If participants chose within 2000 ms, the selected lever lowered, and the image of the corresponding outcome appeared in the lower display (1000 ms). For each response (R_1 and R_2, left or right buttons counterbalanced between participants), the rewarding outcomes were the same O_1 and O_2 rewards used in the Pavlovian learning phase, respectively. In the non-reinforced trials, a non-rewarding outcome ("X") was presented. Crucially, an imbalance in the probability of obtaining the two outcomes was introduced: choosing R_1 (rich option) allowed to obtain O_1 with a 70% probability, whereas choosing R_2 (poor option) allowed to obtain O_2 with a 30% probability. Moreover, a "baiting rule" was inserted (Bari & Gershman, 2023; Rudebeck et al., 2008; Sugrue et al., 2004), such that the longer an agent abstained from choosing a particular option, the greater the probability of receiving a reward for that option. In other words, if an unchosen option would have been rewarded in that trial, the next time that option was selected the reward probability was 100%. Under these circumstances, the probabilistic strategy that allowed to maximize rewards (i.e., the optimal strategy) corresponds to matching: pressing R_1 in 70% of trials and R_2 in 30% of trials (Soltani & Wang, 2006; Sugrue et al., 2004). Moving away from this strategy would constitute suboptimal behavior.

A series of blocks composed of 50 trials were structured to be repeated until a learning criterion was achieved. Each block had an average duration of 3 minutes, followed by multiple-choice questions presented on the display to ensure that all response-outcome associations were correctly established (i.e., "What food did you earn by pressing this lever?", repeated for each lever). The learning criterion consisted of correctly reporting the association between the two responses and the two outcomes two times in a row. If the learning criterion was achieved, the participant moved on to the following phase; otherwise, the task was aborted after four wrong answers (i.e., from a minimum of 2 blocks and a maximum of 8).

Transfer phase

The transfer phase was similar to the instrumental learning phase, with the addition of a component to test the influence of the CSs on decision making. As in the instrumental learning phase, on each trial, an empty slot-machine (**Figure 1c**) was presented (ITI 500-1500 ms), and the appearance of the two levers served as "go-signal" to choose quickly (within 2000 ms). Unlike the instrumental learning phase, however, one of the three CSs previously shown during the Pavlovian learning phase could be presented simultaneously with the appearance of the levers in the upper display of the slot machine.

After each choice, the corresponding outcome appeared on the lower display of the slot machine. Importantly, the probability of obtaining rewards was the same as the instrumental learning phase (i.e., 70% by choosing R_1, 30% by choosing R_2, with the insertion of the same baiting rule), and the CSs were irrelevant for the task. Hence, the optimal decision-making strategy still consisted of pressing R_1 in the 70% of trials and R_2 in the 30% of trials, independently of the CSs. The transfer phase consisted of a single block composed of 30 trials for each CS and 30 trials for the "No cue" condition (i.e., no CSs appear simultaneously with levers, equal to the instrumental learning phase), for a total of 120 trials.

Experimental procedure

Participants were instructed to refrain from eating for 3 hours prior to the experiment. Upon arriving at the laboratory, they signed the informed consent form. Subsequently, two rewarding outcomes were tailored for each participant evaluating their subjective liking for 9 different foods (4 savory foods and 5 sweet foods) on a 5-point Likert scale ranging from 0 (not at all) to 5 (very much). Two foods equally valued were then selected as rewards and the same images were used as outcomes in the experiment. Participants were also asked to rate the current level of hunger from 0 (not at all) to 9 (very much), to ensure a high hunger state (30 Sign-trackers: M = 5.5, SD = 1.9; 30 Goal-trackers: M = 6.6, SD = 1.6). Following this, participants were comfortably seated in a quiet room, positioned centrally in front of a computer screen at 60 cm viewing distance. The foods previously selected were placed on the table to guarantee high motivation throughout the task (Bushong et al., 2010). Then, the eye-tracker was head-mounted on the participant, ensuring a comfortable position and stability during the task through a forehead-chin support placed on the table.

Participants were informed that the number of food pictures visualized during the entire task would be proportional to the amount of food that they would receive after the experiment. They were required to pay attention and follow the instructions reported on the screen. Additionally, they were informed that their responses to the questions presented at the end of each block should be verbalized and selected by the experimenter to ensure the stability of the eye-tracker, which could partially obstruct the keyboard from view.

The entire experimental session lasted approximately 1 hour, the experimenter remained in the room to monitor the participant's performance and check for the correct eye-tracker position throughout the experiment. Furthermore, although the instructions were displayed at the beginning of each phase, the experimenter also provided a brief summary to ensure participants fully understood the task. Before the Pavlovian and instrumental learning phases, four example trials were conducted to further ensure comprehension. At the end of the experiment, participants were provided with at least one of each previously selected reward (i.e., at least two distinct rewards).

Eye-tracking analysis

Eye movements and pupil diameter were recorded with a 2D Eye-Tracker device (Chronos Vision GmbH, Berlin, Germany) at a sampling rate of 200 Hz during the Pavlovian learning phase. The system allows measuring horizontal and vertical movements and consists of two remote cameras that trace the outline of the pupil by means of digital image processing

(online 2D, 11 bit output range) of the eyes that are illuminated by infrared LEDs (940 nm). The measurement range, both horizontal and vertical, is between −40° and +40°, with a resolution of < 0.05° and a measurement error minor than 0.2°. This device delivers binocular data. Since eyes move together under most circumstances, we selected for each participant the signal from the eye that delivered the best accuracy (Carter & Luke, 2020; Hooge et al., 2019), i.e. the eye with the minor number of missing data (left eye: N = 34; right eye: N = 26). The 5-point calibration of the eye-tracker was performed after the Pavlovian learning phase (Hooge et al., 2019). Before the Pavlovian learning phase, participants were instructed to fixate a central fixation cross when no stimuli were presented on the screen (ITI). This ensured correct classification of sign- and goal-trackers by preventing initial shifts of attention towards the sign or goal.

For gaze data analysis, the raw signal from the Pavlovian learning phase was segmented into epochs of 10 seconds centered around the moment of CS appearance (5 seconds of ITI and 5 seconds of CS presentation). The first second after the presentation of the CS was excluded from each trial to exclude the orienting response and pupil light reflex (Gottlieb, 2012; Pietrock et al., 2019). Then, the dwell time during the remaining 4 seconds of the CS presentation was used for both the pupil dilation analysis and the classification in sign- and goal-trackers. Importantly, during these seconds only the CS (and not the corresponding reward) was presented on the screen. Three Areas of Interest (AoI) were considered (see **Figure 1**): two of them were equivalent to the 5 cm upper and lower displays of the slot-machine, respectively reflecting the location of the CS and the location in which the reward would subsequently appear; the third AoI corresponded to the background, i.e., the rest of the screen (Schad et al., 2020). Trials in which participants did not look at the fixation cross during ITI, or at the AoI during CS presentation, were excluded from eye-tracking analysis (number of trials excluded: M = 3.43; SD = 6.04). Dwell time was defined as the amount of time spent within each AoI.

Pupil dilation analysis

Pupil dilation analysis was performed using a customized script running in MATLAB R2024 (The MathWorks, Inc., Natick, MA, USA). Our purpose was to assess only the phasic response of the pupil (Calignano et al., 2024; Mathôt et al., 2018). Pupil dilation in response to CSs typically reflects the acquisition of incentive value by such stimuli (Manohar & Husain, 2015; Schad et al., 2020). We hypothesized that this acquisition should be selective for sign-trackers, as previously demonstrated (Schad et al., 2020).

For data preprocessing, we followed a specific pipeline provided by Calignano and colleagues (2024), who adopted a multiverse approach to pupillometry data analysis to explore the role of the preprocessing phase. We started our data processing by excluding missing data points (M = 0.07%, SD = 0.17), resulting from blink artifacts and invalid data during recording. In addition, participants with more than 30% missing data from the total registration were excluded from further analysis. In this case, all participants were included in subsequent analysis.

Afterward, we selected only the data within the three AoIs described above. Additionally, trials in which data were not recorded for at least 200ms consecutively were discarded (M = 13.3% of total trials, SD = 15.5%) (Mathôt, 2018). As a further step, the trial-by-trial data were baseline-corrected by taking the median pupil value during the 200ms before the

stimulus onset and subtracting this value from the raw data Calignano et al., 2024; Mathôt et al., 2018).
Finally, extreme data, namely those over 2 standard deviations above and below the mean, were excluded (M = 3,8% of total data). Participants with more than 30% extreme values were excluded. As a result, 2 participants were excluded. Thus, the pupil dilation analysis concerns a sample of 58 participants.

Gaze index analysis and sign-trackers vs. goal-trackers categorization

Participants were categorized into sign-trackers and goal-trackers based on the time allocated towards the location of the CS and the location of the reward during the CS presentation in the Pavlovian learning phase. Only the CSs+ trials (Cherkasova et al., 2024; Garofalo & di Pellegrino, 2015; Schad et al., 2020)from the second half of the task (Garofalo & di Pellegrino, 2015) were considered to compute the gaze index.

The gaze index was calculated based on the dwell time spent on the CS location minus the dwell time spent on the reward location, divided by the sum of the dwell time spent on the CS location, the reward location, and the background (Garofalo & di Pellegrino 2015; Schad et al., 2020):

(1) $$Gaze\ index = \frac{Sign - Goal}{Sign + Goal + Background}$$

Therefore, the gaze index was 1 if the entire time was spent looking at the CS, and -1 if the entire time was spent looking at the reward. In other words, the closer to 1 the gaze index, the higher the sign-tracking behavior. Based on this index, the top and bottom 50% of the participants were respectively categorized as sign-trackers and goal-trackers (sign-trackers gaze index: M = 0.88; SD = 0.09; goal-trackers gaze index: M = 0.35; SD = 0.25). Furthermore, to confirm that the gaze index reflected a learned reward-related response rather than a mere attentional bias (i.e. gazing more towards the top vs bottom of the screen (Degni & Garofalo, 2025), the same gaze index was calculated for the CS- and compared to the CS+ index in sign- and goal-trackers.

Computational models

We used data from all three phases of the experiment to fit participants' responses during the transfer phase to various models (see below). Each model's free parameters were fitted to the trial-by-trial responses of individual participants through the maximum posterior estimation method (Gershman, 2016). This procedure involved finding the set of parameters that maximized the likelihood of each participant's trial-by-trial responses, given the model, while being constrained by a regularizing prior. The parameters were estimated using the mfit toolbox (https://github.com/sjgershm/mfit).

i. Pearce-hall associability model. In this model, estimates of associations between CSs, responses, and rewards were updated based on reward prediction error and a dynamic learning rate. More specifically trial-by-trial prediction error (PE) was computed as follows (Li et al., 2011; Pearce & Hall, 1980):

(2) $$PE = O_t - V_t$$

where O_t is the outcome at trial t and V_t is the CS or response associative value at trial t. The PE was then used to update the CS or response associative value:

(3) $$V_{t+1} = V_t + |PE| * PE$$

where V_{t+1} represents the updated CS or response associative value. Here the learning rate is computed as the absolute value of the PE, making it possible to update the learning rate trial-by-trial based on the difference between expected and actual reward.

During the Pavlovian learning phase, the values of the three CSs started from 0 and were updated separately: CS+ $_1$ with O $_1$, CS+ $_2$ with O $_2$, and CS- with nothing. The values were computed based on the sequence of rewards received on a trial-by-trial basis during Pavlovian and instrumental learning phases to obtain the most realistic values at the start of the transfer phase. During the instrumental learning phase, the values of responses R_1 and R_2 were updated similarly, with R $_1$ associated with O $_1$ and R $_2$ associated with O $_2$. In the transfer phase, when a reward was (or was not) obtained, the values of both the response and the CS were updated accordingly. Crucially, in this phase, each CS could be associated with the value of either O_1 or O_2, based on the response provided. Finally, instrumental and the Pavlovian values were combined linearly (Gershman et al., 2021) to obtain an overall value in each trial:

(4) $$WR = (VR * (1 - \omega)) + (VCS_c * \omega)$$

where WR represents the weighted value of the response (computed separately for R $_1$ and R_2), VR the value of the response (R_1 or R_2), VCS_c the value of the current CS for the same outcome of the response (O $_1$ for R $_1$ and O $_2$ for R $_2$), and ω a free parameter which weights CS versus response associative value. In all subsequent RL models, ω operated in the same way as in this equation, with the exception of the "Rescorla–Wagner model with dynamic ω between CS and instrumental value based on baseline arbitration oscillation," where ω was updated according to the specific equations reported in that section. The likelihood was calculated through the softmax function which normalizes stimulus and response values into stochastic choice probability (Luce, 1959) as follows:

(5) $$p(R_c) = \frac{\exp(\beta * WR_c)}{exp(\beta * WR_1) + exp(\beta * WR_2)}$$

where $p(R_c)$ represents the probability of choosing the current response (R $_1$ or R_2), WR_c is the weighted value of the current response (WR_1 or WR_2), WR_1 is the weighted value of R_1, WR_2 is the weighted value of R_2 and β a free parameter standing for the inverse temperature, i.e. the exploitation/exploration balance.

ii. Rescorla-Wagner model. In this model, the estimate of the CS or response associative value was updated as according to the Rescorla-Wagner rule (Wagner & Rescorla, 1972):

(6) $$V_{t+1} = V_t + \alpha * PE$$

Where α is a free parameter for the learning rate, computed separately for CSs (α_{Pav}) and responses (α_{ins}). Thus, in this case, the weight given to PE is considered fixed. The rest of the code is identical to the model we used in the previous section.

iii. Rescorla-Wagner model incorporating coupled learning for response value. This model, which was based on the RL model, associations between responses, and rewards were updated using an additional assumption about the nature of prediction error. More

specifically, this model involved two prediction errors, one for the chosen (PE_c) and one for the unchosen (PE_u) options (Lohrenz et al., 2007):

(7) $PE_c = O_t - VR_{ct}$

(8) $PE_u = (1 - O_t) - VR_{ut}$

Where VR_{ct} is the value at trial t of the chosen option and VR_{ut} is the value at trial t of the unchosen option. Then, response associative values were updated separately:

(9) $VR_{ct+1} = VR_{ct} + \alpha * PE_c$

(10) $VR_{ut+1} = VR_{ut} + \alpha * PE_u$

Other aspects of this model was identical to the standard Rescorla-Wagner model including how CS values were updated.

iv. Rescorla-Wagner model with dynamic arbitration between cs and instrumental value. In this model, we used a set of equations that adaptively adjusted the weight (ω) assigned to the CS associative value relative to the response associative value. At the beginning of the transfer phase, ω was initialized at 0.5, indicating equal weighting between the CS and response values. When participants showed a stronger reliance on the CS associative value (i.e., ω > 0.5), the prediction error for ω was computed as the difference between the obtained outcome (Q) and the current ω. Conversely, when the response associative value was more influential (i.e., ω ≤ 0.5), the prediction error was calculated as the difference between the complement of the outcome ($1 - O_t$) and ω. This rule was implemented because if only strict inequalities (ω > 0.5 and ω < 0.5) were used, then when ω = 0.5 at initialization the update would not occur, leading to ω being stuck at this value. In practice, however, values equal to exactly 0.5 are extremely rare beyond initialization due to the continuous update process, so this adjustment ensures that ω remains dynamically updated throughout the task. This prediction error was then scaled by a learning rate parameter (γ) and used to update ω over trials:

(11) $\omega PE = O_t - \omega_t \quad (if\ \omega > 0.5)$

(12) $\omega PE = (1 - O_t) - \omega_t (if\ \omega < 0.5)$

(13) $\omega_{t+1} = \omega_t + \gamma * \omega PE$

Using this framework, when a participant relied more on the CS associative value and received a reward, ω increased; if no reward was received, ω decreased. The opposite pattern emerged when the response value was more heavily weighted: ω decreased after rewarded outcomes and increased when no reward was delivered. The rationale for this mechanism is to reflect an adaptive arbitration process, allowing the model to flexibly shift control between Pavlovian and instrumental systems based on ongoing feedback and prior reliance. When Pavlovian values dominate, outcomes reinforce or weaken confidence in cue-driven decision-making; when instrumental values dominate, feedback adjusts the weighting in the opposite direction. The remaining components of the model followed the standard Rescorla-Wagner formulation used in the previous section.

v. Rescorla-Wagner model with dynamic arbitration between CS and instrumental value starting with free-parameter. This model used the same equations of the previous model to estimate the arbitration updating based on response and CS associative value during the transfer phase. The only difference is that, in the first trial, ω is computed fitting as free-parameter γ, indicating that the weight of evidence for ω is based on its baseline

value. The rest of the code is identical to the standard Rescorla-Wagner model we used in the previous section.

vi. Rescorla-Wagner model with dynamic ω between CS and instrumental value based on cs and response value reliability starting with free-parameter. This model used the same equations of the previous model to estimate the arbitration updating based on response and CS associative value during the transfer phase. The first was used when the associative value of the CS was lower than the associative value of the response. Using these equations, when a participant considers the CS associative value higher than the response associative value (CS Value > response value) and obtains a reward the value ω increases, while if no reward is obtained it decreases. Vice versa, when a participant considers the response associative value higher than the CS associative value (CS Value < response value) and obtains a reward ω decreases, while if no reward is obtained in this case ω increases. The rest of the code is identical to the Rescorla-Wagner model with dynamic arbitration between CS and instrumental value starting with the free-parameter we used in the previous section.

vii. Rescorla-Wagner model with dynamic ω between CS and instrumental value based on prediction error reliability starting with free-parameter. This model used the same equations of the previous model to estimate the arbitration updating based on response and CS associative value during the transfer phase. The first was used when the absolute value of the prediction error for the CS was lower than the absolute value of the prediction error for the response. Using these equations, when a participant considers the CS more reliable than the response, i.e. when the prediction errors for CS are smaller (|CS PE|<|response PE|), and obtains a reward the value ω increases, while if no reward is obtained it decreases. Vice versa, when a participant considers the response more reliable than the CS (|CS PE|>|response PE|) and obtains a reward ω decreases, while if no reward is obtained in this case ω increases. The rest of the code is identical to the Rescorla-Wagner model with dynamic arbitration between CS and instrumental value starting with the free-parameter we used in the previous section.

viii. Rescorla-Wagner model using different learning rates for every experimental phase. This model used the equations from the standard Rescorla-Wagner model to estimate the CS or response associative value updating. The only difference is that the learning rate was computed separately also for the transfer phase, thus using four parameters for estimating learning rates: one for the Pavlovian learning phase, one for the instrumental learning phase, and two for the transfer phase (one for updating stimulus values and the other for updating instrumental response values). The rest of the code is identical to the standard Rescorla-Wagner model we used in the previous section.

ix. Rescorla-Wagner model differentiating positive and negative prediction errors. In this model, learning rates were allowed to differ for positive and negative PEs, separately for CSs and responses. When PE was positive, one equation governed the associative update, whereas a different one was used when PE was negative, following prior implementations of asymmetrical learning (Daw et al., 2006; Niv et al., 2012):

(14) $V_{t+1} = V_t + \alpha_{pos} * PE(if\ PE > 0)$

(15) $V_{t+1} = V_t + \alpha_{neg} * PE(if\ PE < 0)$

Here, α_{pos} and α_{neg} are free parameters representing the learning rates for positive and negative prediction errors, respectively. These parameters were estimated separately for

CS and response associations. The remaining structure of the model followed the standard Rescorla-Wagner implementation described in the previous section.

x. Rescorla-Wagner model incorporating coupled learning for response value with different learning rates. In this model, separate learning rates were used to update the associative values of chosen and unchosen options. This approach allowed for asymmetric updating depending on whether an option was selected or not during the trial:

(16) $VR_{ct+1} = VR_{ct} + \alpha_{cho} * PE_c$

(17) $VR_{ut+1} = VR_{ut} + \alpha_{unc} * PE_u$

Here, α_{cho} and α_{unc} are free parameters representing the learning rates for the chosen and unchosen options, respectively. These were estimated separately for CS and response values. The remaining aspects of the model followed the structure of the Rescorla-Wagner formulation incorporating coupled learning for response values, as described in the previous section.

xi. Rescorla-Wagner model with dynamic ω between cs and instrumental value based on baseline arbitration oscillation. In this model, arbitration updating during the transfer phase was driven by the difference between the associative values of the CS and the response. The update of the arbitration weight (ω) depended on whether the CS value was higher or lower than the response value on each trial. A passive decay mechanism was also included to bias ω back toward its initial value in the absence of new evidence (Woo et al., 2024):

(18) $\Delta Relt = VCS_c - VR_c$

(19) $\omega_{t+1} = \omega_t + \gamma * \Delta Relt * (1 - \omega_t) + \delta * (\gamma - \omega_t)$ (if $VCS_c > VR_c$)

(20) $\omega_{t+1} = \omega_t + \gamma * |\Delta Relt| * \omega_t + \delta * (\gamma - \omega_t)$ (if $VCS_c < VR_c$)

(21) $WR = (VR * (1 - \omega) * (1 - \rho)) + (VCS_c * \omega * \rho)$

Where γ is the baseline model arbitration rate, or ω at first trial; δ is the passive decay rate that pulls ω toward its initial value γ this additional decay mechanism assumes that ω defaults back to its initial bias in the absence of exogenous input signaling the reliability difference (ΔRelt); ρ is a constant measuring the baseline ratio of signals from the stimulus-based to that from the action-based system independently of time-dependent arbitration weight ω. The rest of the code is identical to the standard Rescorla-Wagner model we used in the previous section.

All models were fitted to the choice data, and the goodness-of-fit was assessed using both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) to identify the model that best captured the behavior. We then used the estimated model parameters from the best-fitting model to investigate differences between sign- and goal-trackers, which were identified based on gaze data alone.

To ensure the reliability of the estimated model parameters, we performed model recovery by simulating 1,000 datasets, each containing the same number of participants as the real dataset and using the same parameter values. We then fitted the model to each simulated dataset to recover the model parameters. Finally, we performed model validation by comparing response rates in the transfer phase to confirm that the best model was able to capture the main aspects of the data.

Statistical analysis

Data were processed offline using custom-made MATLAB scripts (The MathWorks, Inc., Natick, MA, USA) and statistical analyses were performed with RStudio v4.2.1 (RStudio Team, 2016) and Jasp 0.16 (Love et al., 2019) using a Bayesian inferential approach.

The main goals of our study were to test: (a) differences between sign- and goal-trackers in terms of maladaptive Pavlovian bias in decision-making, and (b) differences in the weight of CS associative value versus the CS associative value updating. To this end, we formulated four potential models (or hypotheses) for (a) and three for (b) (the detailed parameters are specified in the "results" section). To directly contrast such hypotheses, Bayesian Informative Hypotheses (BaIn) testing was used (Garofalo et al., 2022, 2024), which enables the comparison of a set of predefined hypotheses through a model selection process, where each model offers a potential explanation for the phenomenon. For each model, the corresponding posterior probability (PMP) was determined using the Bayes theorem and represented as a value between 0 and 1. This value indicates the relative degree of support for each model based on the observed data and the set of competing hypotheses (the sum of all posterior model probabilities adds up to one). The model with the highest PMP represents the hypothesis with the greatest relative likelihood (Gu et al., 2019; Hoijtink, 2012; Hoijtink, Gu, et al., 2019; Hoijtink, Mulder, et al., 2019). To further support model selection, the PMPs can also be compared via Bayes Factor: (a) against the other hypotheses tested, and (b) against its complement hypothesis (H_c), which represents a model containing any set of restrictions between the parameters, except the one represented by the hypothesis being tested.

For all other analyses, Bayesian analyses of variance (ANOVAs) and t-tests were used. The Bayes Factor (BF_{10}) is reported as the probability of the data under the alternative hypothesis (H_1) over the probability of the data under the null hypothesis (H_0), along with its estimated proportional error (err%) (Kruschke, 2021). Bayes factor can also be interpreted in terms of discrete categories of strength of evidence. Following the classification proposed in literature (Andraszewicz et al., 2015; Lee & Wagenmakers, 2013), the BF_{10} can be placed on a continuum from "no evidence" ($BF_{10} = 1$) to "extreme evidence" ($BF_{10} > 100$ for H_1, and $BF_{10} < 1/100$ for H_0), including "anecdotal evidence" ($1 < BF_{10} \leq 3$ for H_1, and $1/3 < BF_{10} < 1$ for H_0), "moderate evidence" ($3 < BF_{10} < 10$ for H_1, and $1/10 < BF_{10} < 1/3$ for H_0), "strong evidence" ($10 < BF_{10} < 30$ for H_1, and $1/30 < BF_{10} < 1/10$ for H_0), "very strong evidence" ($BF_{10} > 100$ for H_1, and $BF_{10} < 1/100$ for H_0). Normality and sphericity assumption of the data were visually inspected and verified assessing values of skewness and kurtosis < 2 for all variables (Hopkins & Weeks, 1990). When assumptions were violated, non-parametric statistics (Bayesian Mann-Whitney test) were used instead, with Kendall's W serving as an estimate of effect size.

Results

Evidence of overall Pavlovian and instrumental learning.

In the Pavlovian learning phase, all participants (N = 60; 100 %) achieved the learning criterion. Specifically, 49 participants (81.6 %; 24 sign-trackers; 25 goal-trackers) answered correctly the CS-outcome associations' question after the minimum two blocks required. 11

participants (18.3 %; 6 sign-trackers and 5 goal-trackers) provided a wrong answer once and repeated the blocks for a total of three times.

In the instrumental learning phase, all participants met the learning criterion after two blocks, as indicated by correctly answering the response-outcome associations question.

Pavlovian learning reflected in CS liking.

We performed a 2×3×2 mixed-measures Bayesian ANOVA with time (pre/post-experiment) and CS ($CS+_1$/$CS+_2$/CS-) as within-subject factors, and group (sign-trackers/goal-trackers) as between-subject factor, using the liking of CSs as dependent variable. Results showed extreme evidence in favor of the alternative hypothesis for the CS by time interaction (BF_{10} = $2.43x10^4$; err% = 1.54) and the main effect of time (BF_{10} = 956.67; err% = 0.99), and very strong evidence for the main effect of CS (BF_{10} = 30.96; err% = 0.73) (**Figure 2a**). All other effects showed moderate (main effect of group: BF_{10} = 0.25; time by group interaction: BF_{10} = 0.21; CS by group interaction: BF_{10} = 0.18) or strong (CS by time by group interaction: BF_{10} = 0.08) evidence in favor of the null hypothesis. Overall, we found a general increase in liking for $CS+_1$ and $CS+_2$ throughout this phase, which does not occur for CS-, confirming the absence of differences between sign-trackers and goal-trackers in terms of behavioral measures of Pavlovian learning.

Gaze data reveal group differences in Pavlovian learning.

We performed a Bayesian independent samples t-test between sign-trackers and goal-trackers using the difference between CS+ and CS- gaze indices as a dependent variable. Results of this analysis provided moderate evidence in favor of the alternative hypothesis (BF_{10} = 8.7; err% < 0.001). Specifically, we found a greater difference in sign-trackers (M = 0.02; SD = 0.09) than in goal-trackers (M = -0.1; SD = 0.07), demonstrating that their classification cannot be attributed to a mere attention bias, but reflects reward-related response (**Figure 2b**). All the eye-tracking results (i.e., gaze index and pupil dilation) refer to the Pavlovian learning phase.

Pupil dilation data reveal differences in Pavlovian learning.

We performed a 3×2 mixed-measures Bayesian ANOVA with CSs ($CS+_1$/$CS+_2$/CS-) as a within-subject factor, and group (sign-trackers/goal-trackers) as a between-subject factor, using the pupil dilation as the dependent variable. This provided strong evidence in favor of the alternative hypothesis for the CS by group interaction (BF_{10} = 19; err% = 1.24), while the main effects of CS (BF_{10} = 0.06; err% = 0.83) and group (BF_{10} = 0.50; err% = 2.69) showed respectively strong and anecdotal evidence in favor of the null hypothesis. Overall, we found that sign-trackers discriminated at the psychophysiological level between CSs+ and CS-, in contrast, goal-trackers did not present such an effect (**Figure 2c**). Thus, in line with the literature, results suggest a greater incentive value attribution to CSs+ in sign-trackers only (**Figure 2c**).

Similar Instrumental learning by sign- and goal-trackers.

To confirm the acquisition of the optimal decision-making strategy and examine possible differences between sign-trackers and goal-trackers during the instrumental learning phase, we performed a 2×2 mixed-measures Bayesian ANOVA with responses (R_1 /R_2) as a within-subject factor and group (sign-trackers/goal-trackers) as a between-subject factor, using the relative percentage R_1 and R_2 (over the total) as dependent variable. This analysis provided extreme evidence in favor of the alternative hypothesis for the main effect of response (BF_{10} = 2.85x10^{17}; err% = 1.257), while the main effect of group (BF_{10} = 0.206; err% = 0.971) and the response by group interaction (BF_{10} = 0.473; err% = 1.273) showed respectively moderate and anecdotal evidence in favor of the null hypothesis (**Figure 2d**). Overall, results demonstrate the acquisition of the optimal decision-making strategy by both sign-trackers and goal-trackers, which consisted of pressing more R_1 than R_2.

Different Pavlovian bias by sign- and goal-trackers during the transfer phase.

To identify possible differences in maladaptive Pavlovian bias in decision-making between sign-trackers and goal-trackers, we examined performance during the transfer phase by testing Bayesian Informative Hypotheses on the percentage of R_1 responses (i.e., the rich option) across the two CSs. Pavlovian bias corresponds to: (a) an increased percentage of R_1 responses in the presence of the CS+ $_1$, and (b) a reduced percentage of R_1 responses in the presence of the CS+ $_2$, compared to the two control conditions (CS- and No cue). Specifically, we formulated four hypotheses to be directly compared.

The first hypothesis posited that only sign-trackers exhibit Pavlovian bias:

$$H_1: [CS+_1 > (CS- = \text{No cue}) > CS+_2]_{\text{sign-trackers}} \;\&\; [CS+_1 = CS- = \text{No cue} = CS+_2]_{\text{goal-trackers}}$$

The second hypothesis posited that only goal-trackers exhibit Pavlovian bias:

$$H_2: [CS+_1 = CS- = \text{No cue} = CS+_2]_{\text{sign-trackers}} \;\&\; [CS+_1 > (CS- = \text{No cue}) > CS+_2]_{\text{goal-trackers}}$$

The third hypothesis posited that both groups exhibit Pavlovian bias:

$$H_3: [CS+_1 > (CS- = \text{No cue}) > CS+_2]_{\text{sign-trackers}} \;\&\; [CS+_1 > (CS- = \text{No cue}) > CS+_2]_{\text{goal-trackers}}$$

The fourth hypothesis posited that no group exhibits Pavlovian bias:

$$H_4: [CS+_1 = CS- = \text{No cue} = CS+_2]_{\text{sign-trackers}} \;\&\; [CS+_1 = CS- = \text{No cue} = CS+_2]_{\text{goal-trackers}}$$

Results revealed that H_1 has the highest relative posterior probability, indicating it as the strongest hypothesis, both when excluding and including the complement hypothesis H_c (**Figure 3a**; **Table 2**). Consistently, H_1 also had the highest Bayes Factor computed relative to H_2 (BF_{12} = 40.37), H_3 (BF_{13} = 3.37), and H_4 (BF_{14} = 10.48). Visual inspection of the data confirms that, in the CS+ $_1$ and control (CS- and No cue) conditions, both groups maintain the optimal decision-making strategy, consisting of choosing more R_1 than R_2 (**Figure 3b**).

Crucially, the main differences between the two groups emerge in the presence of the CS+ $_2$: while goal-trackers exhibited a higher selection of the correct response ($R_1 > R_2$), sign-trackers showed the maladaptive Pavlovian bias, as evident from an increased percentage of R_2 responses compared to the control conditions.

The trial-by-trial average response over time further revealed that at the beginning of the transfer phase, both groups were affected by CS+ $_2$ (increased percentage of R $_2$) (**Figure 3c**). Subsequently, while goal-trackers rapidly adapted their choices and returned to the optimal decision-making strategy, sign-trackers were not able to overcome the Pavlovian bias, as evidenced by the overlapping confidence intervals for R_1 and R_2 responses (**Figure 3b,c**). To further explore this observation, a Bayesian generalized linear mixed model (binomial family, logit link) was fitted to predict participants' responses (R_1/R_2) as a function of group (GT vs ST) and trial (1–30) during CS+ $_2$ trials (Barr et al., 2013). The model used the default auto-scaled weakly informative prior (rstanarm package). Random effects included a random intercept for participants and a random slope for trials (1 + Trial | Subject). The posterior mean of the standard deviation for the two random parameters indicated highly variable baseline response levels between participants (σIntercept = 1.087) and low variability between trials (σTrial = 0.052). Diagnostic checks indicated satisfactory convergence and mixing, with the $\hat{R}$ statistic for all fixed-effect parameters equaling 1.00. The group by trial interaction revealed that the probability of responding R_1 increased credibly for GTs (β = 0.043, 95% CI [0.016, 0.071]; credible interval did not include 0), whereas no clear trend was observed for STs (β = 0.017, 95% CI [–0.011, 0.044]; credible interval included 0). The difference in slopes ($\Delta\beta \approx 0.026$) suggests that GT participants showed a steeper increase in responding over time compared with STs. Overall, the model indicates that GTs were more likely to produce a response and that their performance improved more consistently across trials, whereas STs displayed a flatter learning curve with limited change in response probability over time.

Computational modeling reveals interaction between Pavlovian and instrumental learning in both groups.

To identify possible mechanisms underlying the differences in Pavlovian bias exhibited by participants during the transfer phase, we fitted the choice data on a trial-by-trial basis using various computational models (see Methods for more details). To determine the best-fitting model, we used the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) as measures of goodness-of-fit (Vrieze, 2012).

We found that the best model was the Rescorla-Wagner model with dynamic arbitration between CS and instrumental value, with an initial free parameter for the arbitration weight (**Figure 4a,b**; **Table 1**). Additionally, all other models involving dynamic interaction between Pavlovian and instrumental learning showed a goodness-of-fit close to the best-fitting model. Importantly, we performed model recovery and model validation to ensure that the parameters of the best-fitting model could be accurately estimated and that this model captures the main aspects of the behavioral data. For model recovery, we found a significant correlation between the parameters from the real and simulated data (bootstrapping; **Figure 4c**), demonstrating that we were able to estimate the underlying model parameters with

good accuracy. For model validation, we compared the response rates of simulated and actual data in the transfer phase and found similar patterns (**Figure 4d**).

Overall, these results suggest that during the transfer phase, the choice behavior of most participants was influenced by both Pavlovian cue values learned during the Pavlovian learning phase and action values learned during the instrumental learning phase, while these values were updated during the transfer phase.

Slower Pavlovian update in sign-trackers accounts for differences between sign- and goal-trackers.

To determine the possible mechanisms underlying the observed differences in Pavlovian bias between sign- and goal-trackers, we compared the estimated model parameters from the best-fitting model for the two groups. Specifically, we examined two parameters crucial for testing our hypotheses: the Pavlovian learning rate (α_{Pav}) and the γ parameter, which denotes the baseline weighting of Pavlovian cue values relative to instrumental action values during decision-making in the transfer phase. The Pavlovian learning rate captures the speed at which the associative value of CSs is updated based on prediction error, with lower values corresponding to slower or more rigid updating of cue values. In contrast, γ captures the default influence of Pavlovian CS on instrumental behavior, independent of learned action-outcome contingencies, with higher values indicating a greater reliance on Pavlovian cue values.

More specifically, we tested four hypotheses using Bayesian Informative Hypotheses testing. The first hypothesis posited that sign-trackers exhibit a lower α_{Pav} compared to goal-trackers, while the two groups would not differ on γ, indicating that the effect is driven by a slower CS associative value updating in sign-trackers:

$$H_1: [\alpha_{Pav}]_{sign\text{-}trackers} < [\alpha_{Pav}]_{goal\text{-}trackers} \;\&\; [\gamma]_{sign\text{-}trackers} = [\gamma]_{goal\text{-}trackers}$$

The second hypothesis posited no differences in α_{Pav} between groups, but higher γ in sign-trackers compared to goal-trackers, indicating that the effect is driven by an increased weight of the CS associative value on instrumental values in sign-trackers:

$$H_2: [\alpha_{Pav}]_{sign\text{-}trackers} = [\alpha_{Pav}]_{goal\text{-}trackers} \;\&\; [\gamma]_{sign\text{-}trackers} > [\gamma]_{goal\text{-}trackers}$$

The third hypothesis posited that sign-trackers would exhibit both a lower α_{Pav} and a higher γ compared to goal-trackers:

$$H_3: [\alpha_{Pav}]_{sign\text{-}trackers} < [\alpha_{Pav}]_{goal\text{-}trackers} \;\&\; [\gamma]_{sign\text{-}trackers} > [\gamma]_{goal\text{-}trackers}$$

The fourth hypothesis posited no differences between groups in both α_{Pav} and γ:

$$H_4: [\alpha_{Pav}]_{sign\text{-}trackers} = [\alpha_{Pav}]_{goal\text{-}trackers} \;\&\; [\gamma]_{sign\text{-}trackers} = [\gamma]_{goal\text{-}trackers}$$

H_1 emerged as the strongest hypothesis, demonstrating the highest relative posterior model probability, even when accounting for H_c (**Figure 5a**; **Table 3**). Additionally, H_1 showed the highest Bayes Factor when compared to H_2(BF_{12} = 95.344), H_3 (BF_{13} = 8.271), and H_4 (BF_{14} = 10.948). This pattern is also evident when comparing α_{Pav} and γ values for sign- and goal-trackers (**Figure 5b**). Model simulations based on the estimated model parameters further

revealed differences between the two groups, consistent with slower Pavlovian updates in sign-trackers. More specifically, for sign-trackers, the estimated values of $CS+_2$ for O_1 and O_2 for O1 and O2 begin to overlap after about 10 presentations of the CS, as indicated by the convergence of their 95% confidence intervals (**Figure 5c**). In contrast, this overlap emerges much earlier for goal-trackers, after approximately 5 CS presentations, reflecting faster updating of outcome-specific values. This pattern is also evident for CS- values, which stop overlapping after about 13 presentations for sign-trackers and after 5 for goal-trackers. Moreover, for $CS+_1$, goal-trackers exhibit a steeper increase in the associative value for O2, suggesting a more rapid updating of the CS–outcome association in this group.

There were no a priori hypotheses regarding the learning rate for response value update (α_{ins}) and the inverse temperature (β, i.e., the trade-off between exploitation and exploration), or the average effective arbitration weight across trials (ω_{mean}). However, as an exploratory analysis, we tested whether these parameters differed for sign- and goal-trackers. The rationale for including them is that these indices provide complementary information on potential group differences in decision processes: α_{ins} captures updating of instrumental values, β reflects choice stochasticity, and ω_{mean} summarizes the effective arbitration weight applied across transfer trials by integrating both baseline bias and trial-by-trial updating. The results indicated no significant differences between the two groups for none of these parameters (β: BF_{10} = 0.279; W = 458 | α_{ins}: BF_{10} = 0.307; err% = 0.010 | ω_{mean}: BF_{10} = 1.072; err% = 0.009).

Discussion

Using a combination of experimental and computational approaches, here we identified two distinct patterns of Pavlovian bias between the sign-trackers, who approach reward-predictive cues, and goal-trackers, who approach the location of reward delivery (Colaizzi et al., 2020; Heck et al., 2024). Experimentally, we used a three-stage Pavlovian-to-Instrumental Transfer task to test how Pavlovian cues bias participants' choices toward their associated outcomes, even when these cues were task-irrelevant, during the transfer phase. Computationally, we developed and tested various mechanistic models to evaluate their ability to capture choice data and identify potential differences between sign- and goal-trackers.

Consistent with our hypothesis, sign-trackers exhibited a stronger Pavlovian bias than goal-trackers. Specifically, during the transfer phase, only sign-trackers increased their selection of the poor option (R_2) when presented with the associated reward-predictive cue ($CS+_2$). In contrast, goal-trackers maintained their decision-making strategy from the instrumental learning phase, indicating they were more able to disregard the Pavlovian cues when these were no longer relevant to reward delivery.

Importantly, both groups successfully learned the optimal (intended as reward gain maximization) decision-making strategy during the instrumental learning phase. They also maintained this strategy in all control conditions during the transfer phase, namely, when presented with the cue associated with the rich option ($CS+_1$), the non-predictive cue (CS−), or when no cue was presented (No Cue). This indicates that the deviation observed in sign-trackers during the $CS+_2$ condition cannot be explained by a failure to learn or retrieve the

optimal strategy. Moreover, both groups showed comparable levels of Pavlovian learning, as they accurately reported CS–outcome associations and rated CS+ cues more favorably. These results suggest that the observed behavioral bias is not due to differences in associative learning per se.

In line with previous work (Schad et al., 2020), we observed group differences in terms of psychophysiological responses during Pavlovian learning: only sign-trackers displayed greater pupil dilation to reward predictive cues (CS+s) compared to neutral cues (CS-). Given that pupil dilation reliably reflects the incentive salience attributed to Pavlovian cues (Finke et al., 2021; Pietrock et al., 2019), this finding supports the view that sign-trackers assign greater motivational value to CSs (Felix & Flagel, 2024; Robinson & Flagel, 2009). Moreover, this physiological marker based on pupil dilation provides independent evidence for the validity of our sign- vs goal-tracker classification based on gaze fixation data. Critically, the selective acquisition of heightened incentive value in sign-trackers may underlie their increased susceptibility to maladaptive Pavlovian bias. In this group, cues can act as “motivational magnets” that capture attention (Le Pelley et al., 2016), which can, in turn, interfere with the optimal decision-making strategy (Watson et al., 2014) by reactivating instrumental responses previously paired with those outcomes (Finotti et al., 2025).

Interestingly, a more compelling pattern emerges when examining behavior over time (**Figure 5c**). In the early trials of the transfer phase (~5-8 initial trials), both sign- and goal-trackers tended to favor the poor option (R_2) when presented with its associated cue ($CS+_2$). This suggests that both groups initially interpreted the cue as potentially informative (Dorfman & Gershman, 2019), a strategy consistent with previous findings showing that participants often expect cues to guide reward acquisition (Cartoni et al., 2013). However, as the task progresses and feedback reveals that the cues are not predictive of choice outcomes, participants should mainly rely on instrumental values learned earlier. Here, we found a critical group difference: goal-trackers rapidly adapted and returned to use of those values, whereas sign-trackers persisted in their maladaptive bias, showing a markedly slower adjustment. This inflexibility is unique to the $CS+_2$ condition and is absent for all other control conditions ($CS+_1$, CS−, and No Cue).

Reduction of the impact of Pavlovian cues on decision making can happen in at least two non-exclusive ways: by reevaluating (or updating) the predictive value of cues, and by reducing the influence of cue values on choice. Our computational model results provide the first direct evidence that individual differences in Pavlovian learning styles are linked to an inflexibility in cue value updates rather than static overvaluation of the Pavlovian cue relative to instrumental actions. Specifically, sign-trackers exhibited a slower Pavlovian learning rate than goal-trackers, while showing no difference in the weighting of Pavlovian vs. instrumental values. Consequently, sign-trackers continued to rely on outdated cue values, leading to persistent suboptimal choices even when cues are no longer relevant. This behavioral rigidity aligns with prior findings that sign-trackers are more resistant to extinction than goal-trackers (Ahrens et al., 2016), suggesting a habitual, reward-independent behavioral pattern.

Strikingly, this pattern mirrors mechanisms implicated in addiction, where no-longer predictive cues continue to drive behavior despite adverse consequences (Redish, 2004). Thus, our findings bridge insight from animal learning theories with human computational psychiatry and highlight an impaired cue-updating process in the new context as a potential

diagnostic marker of maladaptive behavior in response to changing environmental contingencies.

It is important to note that the present task implements an outcome-specific PIT design, whereas previous studies reporting enhanced PIT effects in sign-trackers (Garofalo & di Pellegrino, 2015; Schad et al., 2020) primarily assessed general PIT. These two forms of PIT rely on only partially overlapping neural and cognitive mechanisms (Finotti et al., 2025; Garofalo et al., 2021; Prévost et al., 2012), making it difficult to establish whether the mechanism we identified —namely, reduced updating of Pavlovian cue values —represents a general feature of Pavlovian control or a process specific to outcome-dependent learning. Another important difference with previous literature is the presence of reward delivery during the transfer phase. Although it may appear that only our design allows for potential re-learning of cue–outcome associations, even classical PIT paradigms conducted under extinction can also be viewed as involving new learning. Indeed, exposure to Pavlovian cues without reinforcement is known to induce extinction learning that reduces or updates the acquired value of conditioned stimuli (Domjan, 2005; Pavlov, 1927). From this perspective, the enhanced general PIT previously reported in sign-trackers could plausibly reflect a slower extinction –or, equivalently, slower updating—of Pavlovian values, a possibility consistent with evidence from animal models (Ahrens et al., 2016) and broadly compatible with our present findings. However, since general and outcome-specific PIT engage distinct processes, it remains an open question whether these effects stem from a common underlying mechanism involving Pavlovian value updating or from separate ones. Future studies will be necessary to resolve this issue.

Taken together, our findings refine the understanding of interindividual differences in decision-making by showing that vulnerability to Pavlovian bias may stem not from how Pavlovian associations are formed, but from how flexibly they are updated over time. This reframes the problem from cue reactivity to the temporal dynamics of cue-reward learning, offering a novel computational perspective for early risk detection and targeted interventions. Our study provides a direct experimental link between sign-tracking and suboptimal behavior in humans, marking a crucial step toward translational applications.

Future research could extend this framework to clinical populations suffering from psychiatric disorders (Felix & Flagel, 2024), exploring whether learning flexibility can be enhanced through pharmacologic (Coddington et al., 2023; Grossman et al., 2022), transcranial (Biernacki et al., 2023; Ke et al., 2019), or cognitive-behavioral (Dercon et al., 2024; Pittig et al., 2023) interventions aimed at modulating learning rates. Embedding cue value inflexibility into predictive models of risk may enable personalized, mechanistically grounded psychiatric care. By identifying a novel computational signature of maladaptive cue reactivity, this work offers a potential biomarker for disorders characterized by disrupted Pavlovian control, such as addiction and impulse-control conditions. Bridging psychophysiology, behavioral neuroscience, and computational psychiatry, our findings lay the ground for targeted, early interventions based on individual profiles of cue sensitivity and learning dynamics.

Acknowledgments

We would like to thank Benedetta Luciani and Luca Pascucci for their invaluable help in data collection, Dr. Yoann Stussi for fruitful discussions on computational modeling, and Dr. Giulia Calignano for her assistance with pupil dilation analyses.

Data availability

The task, datasets, and code used in the paper are freely available on the following Open Science Framework page https://osf.io/nkm2e/, following FAIR (findable, accessible, interoperable, and reusable) data principles (Wilkinson et al., 2016).

Funding

The authors of this work are supported by the following grants:

- European Union's NextGenerationEU and the Italian Ministry of Research (MUR) National Recovery and Resilience Plan (NRRP-PNRR) Progetti di Rilevante Interesse Nazionale (PRIN) 2022 grant, project titled “Individual variations in resisting temptation: the balance between motivational and inhibitory control” (ES PNRR - M4C2 - I1.3 - Prot. P2022KAZ45_001 – CUP J53D23017240001).

- European Union's NextGenerationEU and the Italian Ministry of Research (MUR) National Recovery and Resilience Plan (NRRP-PNRR) Progetti di Rilevante Interesse Nazionale (PRIN) 2022 grant, project titled “Decision-making through the lens of the motor system: a neurocomputational perspective” (Prot. 2022W5TTK8 - CUP J53D23019450006).

- European Union's NextGenerationEU and the Italian Ministry of Research (MUR) National Recovery and Resilience Plan (NRRP-PNRR), project titled “MNESYS” (N. PE0000006; DN. 1553 11.10.2022)

The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Author contributions

Luigi A.E. Degni: conceptualization, methodology, software, formal analysis, data curation, Writing - Original Draft, visualization; **Lorenzo Mattioni**: conceptualization, formal analysis, data curation, Writing - Original Draft, visualization; **Claudio Danti**: validation, formal analysis, investigation, Writing - Review & Editing; **Valentina Bernardi**: investigation, Writing - Review & Editing; **Gianluca Finotti**: conceptualization, methodology, software, Writing - Review & Editing; **Marco Badioli**: validation, Writing - Review & Editing; **Francesca Starita**: Writing - Review & Editing, funding acquisition; **Alireza Soltani**: visualization, supervision, funding acquisition, Writing - Review & Editing; **Giuseppe di Pellegrino**: conceptualization, methodology, visualization, supervision, project administration, funding acquisition, Writing - Review & Editing; **Sara Garofalo**:

903 conceptualization, methodology, visualization, software, formal analysis, supervision,
904 project administration, funding acquisition, Writing - Review & Editing.

905

References

Ahrens, A. M., Singer, B. F., Fitzpatrick, C. J., Morrow, J. D., & Robinson, T. E. (2016). Rats that sign-track are resistant to Pavlovian but not instrumental extinction. *Behavioural Brain Research*, *296*, 418–430. https://doi.org/10.1016/J.BBR.2015.07.055

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2015). An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*, *41*(2), 521–543. https://doi.org/10.1177/0149206314560412

Badioli, M., Degni, L. A. E., Dalbagno, D., Danti, C., Starita, F., di Pellegrino, G., Benassi, M., & Garofalo, S. (2024). Unraveling the influence of Pavlovian cues on decision-making: A pre-registered meta-analysis on Pavlovian-to-instrumental transfer. *Neuroscience & Biobehavioral Reviews*, *164*, 105829. https://doi.org/10.1016/J.NEUBIOREV.2024.105829

Bari, B. A., & Gershman, S. J. (2023). Undermatching Is a Consequence of Policy Compression. *Journal of Neuroscience*, *43*(3), 447–457. https://doi.org/10.1523/JNEUROSCI.1003-22.2022

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/J.JML.2012.11.001

Berridge, K. C. (2000). Reward learning: Reinforcement, incentives, and expectations. *Psychology of Learning and Motivation - Advances in Research and Theory*, *40*, 223–278. https://doi.org/10.1016/S0079-7421(00)80022-5

Biernacki, K., Myers, C. E., Cole, S., Cavanagh, J. F., & Baker, T. E. (2023). Prefrontal transcranial magnetic stimulation boosts response vigour during reinforcement learning in healthy adults. *European Journal of Neuroscience*, *57*(4), 680–691. https://doi.org/10.1111/EJN.15905

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*(11), 681–684. https://doi.org/10.1037/H0040090

Brown, P. L., & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior*, *11*(1), 1–8. https://doi.org/10.1901/JEAB.1968.11-1

Bushong, B., King, L. M., Camerer, C. F., & Rangel, A. (2010). Pavlovian Processes in Consumer Choice: The Physical Presence of a Good Increases Willingness-to-Pay. *American Economic Review*, *100*(4), 1556–1571. https://doi.org/10.1257/AER.100.4.1556

Calignano, G., Girardi, P., & Altoè, G. (2024). First steps into the pupillometry multiverse of developmental science. *Behavior Research Methods*, *56*(4), 3346–3365. https://doi.org/10.3758/S13428-023-02172-8/FIGURES/10

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/S13428-012-0186-0/FIGURES/3

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62. https://doi.org/10.1016/J.IJPSYCHO.2020.05.010

Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental Transfer: A review. *Neuroscience & Biobehavioral Reviews*, *71*, 829–848. https://doi.org/10.1016/J.NEUBIOREV.2016.09.020

Cartoni, E., Puglisi-Allegra, S., & Baldassarre, G. (2013). The three principles of action: A Pavlovian-instrumental transfer hypothesis. *Frontiers in Behavioral Neuroscience*, *0*(NOV), 153. https://doi.org/10.3389/FNBEH.2013.00153/BIBTEX

Cherkasova, M. V., Clark, L., Barton, J. J. S., Stoessl, A. J., & Winstanley, C. A. (2024). Risk-promoting effects of reward-paired cues in human sign- and goal-trackers. *Behavioural Brain Research*, *461*, 114865. https://doi.org/10.1016/J.BBR.2024.114865

Coddington, L. T., Lindo, S. E., & Dudman, J. T. (2023). Mesolimbic dopamine adapts the rate of learning from action. *Nature 2023 614:7947*, *614*(7947), 294–302. https://doi.org/10.1038/s41586-022-05614-z

Colaizzi, J. M., Flagel, S. B., Joyner, M. A., Gearhardt, A. N., Stewart, J. L., & Paulus, M. P. (2020). Mapping sign-tracking and goal-tracking onto human behaviors. *Neuroscience & Biobehavioral Reviews*, *111*, 84–94. https://doi.org/10.1016/J.NEUBIOREV.2020.01.018

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/J.NEURON.2011.02.027/ATTACHMENT/62B9719B-7BB0-493C-A1F6-D539039814CE/MMC1.PDF

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095). https://doi.org/10.1038/nature04766

Degni, L. A. E., Dalbagno, D., Starita, F., Benassi, M., di Pellegrino, G., & Garofalo, S. (2022). General Pavlovian-to-instrumental transfer in humans: Evidence from Bayesian inference. *Frontiers in Behavioral Neuroscience*, *16*. https://doi.org/10.3389/FNBEH.2022.945503

Degni, L. A. E., Danti, C., Finotti, G., Starita, F., di Pellegrino, G., & Garofalo, S. (2025). Pavlovian bias instigates suboptimal choices in humans. *Behaviour Research and Therapy*, *195*, 104906. https://doi.org/10.1016/J.BRAT.2025.104906

Degni, L. A. E., & Garofalo, S. (2025). Toward a Translational Model of Sex-Associated Pavlovian Phenotypes. *Addiction Biology*, *30*(6), e70054. https://doi.org/10.1111/ADB.70054

Degni, L. A. E., Garofalo, S., Finotti, G., Starita, F., Robbins, T. W., & di Pellegrino, G. (2024). Sex differences in motivational biases over instrumental actions. *Npj Science of Learning*, *9*(1), 1–13. https://doi.org/10.1038/s41539-024-00246-6

Dercon, Q., Mehrhof, S. Z., Sandhu, T. R., Hitchcock, C., Lawson, R. P., Pizzagalli, D. A., Dalgleish, T., & Nord, C. L. (2024). A core component of psychological therapy causes adaptive changes in computational learning mechanisms. *Psychological Medicine*, *54*(2), 327–337. https://doi.org/10.1017/S0033291723001587

Domjan, M. (2005). Pavlovian conditioning: A functional perspective. *Annual Review of Psychology*, *56*(Volume 56, 2005), 179–206. https://doi.org/10.1146/ANNUREV.PSYCH.55.090902.141409/CITE/REFWORKS

Dorfman, H. M., & Gershman, S. J. (2019). Controllability governs the balance between Pavlovian and instrumental action selection. *Nature Communications*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-13737-7

Duckworth, J. J., Wright, H., Christiansen, P., Rose, A. K., & Fallon, N. (2022). Sign-tracking modulates reward-related neural activation to reward cues, but not reward feedback. *European Journal of Neuroscience*, *56*(7), 5000–5013. https://doi.org/10.1111/EJN.15787

Felix, P. C., & Flagel, S. B. (2024). Leveraging Individual Differences in Cue-Reward Learning to Investigate the Psychological and Neural Basis of Shared Psychiatric Symptomatology: The Sign-Tracker/Goal-Tracker Model. *Behavioral Neuroscience*. https://doi.org/10.1037/bne0000590

Finke, J. B., Roesmann, K., Stalder, T., & Klucken, T. (2021). Pupil dilation as an index of Pavlovian conditioning. A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *130*, 351–368. https://doi.org/10.1016/J.NEUBIOREV.2021.09.005

Finotti, G., Degni, L. A. E., Badioli, M., Dalbagno, D., Starita, F., Bardi, L., Huang, Y., Wei, J., Sirigu, A., Gazzola, V., Pellegrino, G. di, & Garofalo, S. (2025). Cortical Beta Power Reflects the Influence of Pavlovian Cues on Human Decision-Making. *Journal of Neuroscience*, *45*(6). https://doi.org/10.1523/JNEUROSCI.0414-24.2024

Flagel, S. B., Akil, H., & Robinson, T. E. (2009). Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. *Neuropharmacology*, *56*(SUPPL. 1), 139–148. https://doi.org/10.1016/j.neuropharm.2008.06.027

Garofalo, S., Battaglia, S., & di Pellegrino, G. (2019). Individual differences in working memory capacity and cue-guided behavior in humans. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-43860-W

Garofalo, S., Battaglia, S., Starita, F., & di Pellegrino, G. (2021). Modulation of cue-guided choices by transcranial direct current stimulation. *Cortex*, *137*, 124–137. https://doi.org/10.1016/j.cortex.2021.01.004

Garofalo, S., & di Pellegrino, G. (2015). Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. *Frontiers in Behavioral Neuroscience*, *9*(JUNE), 163. https://doi.org/10.3389/FNBEH.2015.00163/BIBTEX

Garofalo, S., Finotti, G., Orsoni, M., Giovagnoli, S., & Benassi, M. (2024). Testing Bayesian Informative Hypotheses in 5 steps with R and JASP. *Advances in Methods and Practices in Psychological Science*.

Garofalo, S., Giovagnoli, S., Orsoni, M., Starita, F., & Benassi, M. (2022). Interaction effect: Are you doing the right thing? *PLOS ONE*, *17*(7), e0271668. https://doi.org/10.1371/JOURNAL.PONE.0271668

Garofalo, S., & Robbins, T. W. (2017). Triggering avoidance: dissociable influences of aversive pavlovian conditioned stimuli on human instrumental behavior. *Frontiers in Behavioral Neuroscience*, *11*, 63. https://doi.org/10.3389/FNBEH.2017.00063/BIBTEX

Garofalo, S., Sagliano, L., Starita, F., Trojano, L., & di Pellegrino, G. (2020). Subliminal determinants of cue-guided choice. *Scientific Reports 2020 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-68926-y

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, *71*. https://doi.org/10.1016/j.jmp.2016.01.006

Gershman, S. J., Guitart-Masip, M., & Cavanagh, J. F. (2021). Neural signatures of arbitration between Pavlovian and instrumental action selection. *PLOS Computational Biology*, *17*(2), e1008553. https://doi.org/10.1371/JOURNAL.PCBI.1008553

Gottlieb, J. (2012). Attention, Learning, and the Value of Information. *Neuron*, *76*(2), 281–295. https://doi.org/10.1016/J.NEURON.2012.09.034/ASSET/6091399B-A0B3-4606-B1EF-9FD54183F010/MAIN.ASSETS/GR5.JPG

Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2022). Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, *32*(3), 586-599.e7. https://doi.org/10.1016/J.CUB.2021.12.006/ASSET/E750A96D-8AB7-4317-A073-467C63BE2CDC/MAIN.ASSETS/GR7.JPG

Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: a program for Bayesian testing of order constrained hypotheses in structural equation models. *Https://Doi.Org/10.1080/00949655.2019.1590574*, *89*(8), 1526–1553. https://doi.org/10.1080/00949655.2019.1590574

Hearst, E., & Jenkins, M. H. (1974). Sign-tracking : the stimulus-reinforcer relation and directed action. *Psychonomic Society*. https://cir.nii.ac.jp/crid/1130000798379464448

Heck, M., Durieux, N., Anselme, P., & Quertemont, E. (2024). Implementations of sign- and goal-tracking behavior in humans: A scoping review. *Cognitive, Affective and Behavioral Neuroscience*, 1–28. https://doi.org/10.3758/S13415-024-01230-8/FIGURES/5

Hoijtink, H. (2012). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Chapman & Hall/CRC.

Hoijtink, H., Gu, X., & Mulder, J. (2019). Bayesian evaluation of informative hypotheses for multiple populations. *British Journal of Mathematical and Statistical Psychology*, *72*(2), 219–243. https://doi.org/10.1111/bmsp.12145

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A Tutorial on Testing Hypotheses Using the Bayes Factor. *Psychological Methods*. https://doi.org/10.1037/met0000201

Hooge, I. T. C., Holleman, G. A., Haukes, N. C., & Hessels, R. S. (2019). Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods*, *51*(6), 2712–2721. https://doi.org/10.3758/S13428-018-1135-3/FIGURES/7

Hopkins, K. D., & Weeks, D. L. (1990). Tests for Normality and Measures of Skewness and Kurtosis: Their Place in Research Reporting. *Http://Dx.Doi.Org/10.1177/0013164490504001*, *50*(4), 717–729. https://doi.org/10.1177/0013164490504001

Ke, Y., Wang, N., Du, J., Kong, L., Liu, S., Xu, M., An, X., & Ming, D. (2019). The effects of transcranial direct current stimulation (tDCS) on working memory training in healthy young adults. *Frontiers in Human Neuroscience*, *13*, 421402. https://doi.org/10.3389/FNHUM.2019.00019/BIBTEX

Kruschke, J. K. (2021). Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, *5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV), 62627. https://doi.org/10.3389/FPSYG.2013.00863/ABSTRACT

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, *84*(3), 555–579. https://doi.org/10.1901/JEAB.2005.110-04

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000064

Lee, M. D., & Wagenmakers, E. J. (2013). Bayesian cognitive modeling: A practical course. In *Cambridge University Press*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10). https://doi.org/10.1038/nn.2904

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), 9493–9498. https://doi.org/10.1073/PNAS.0608842104/SUPPL_FILE/IMAGE916.GIF

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, *88*(1), 1–17. https://doi.org/10.18637/JSS.V088.I02

Luce, R. D. (1959). *Individual choice behavior* (Vol. 4). Wiley.

Manohar, S. G., & Husain, M. (2015). Reduced pupillary reward sensitivity in Parkinson's disease. *Npj Parkinson's Disease 2015 1:1*, *1*(1), 1–4. https://doi.org/10.1038/npjparkd.2015.26

Marzuki, A. A., Banca, P., Garofalo, S., Degni, L. A. E., Dalbagno, D., Badioli, M., Sule, A., Kaser, M., Conway-Morris, A., Sahakian, B. J., & Robbins, T. W. (2024). Compulsive avoidance in youths and adults with OCD: an aversive pavlovian-to-instrumental transfer study. *Translational Psychiatry*, *14*(1), 1–12. https://doi.org/10.1038/s41398-024-03028-1

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/JOC.18

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/S13428-017-1007-2/FIGURES/6

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314. https://doi.org/10.3758/S13428-011-0168-7

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*(2). https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Ostlund, S. B., & Marshall, A. T. (2021). Probing the role of reward expectancy in Pavlovian-instrumental transfer. *Current Opinion in Behavioral Sciences*, *41*, 106–113. https://doi.org/10.1016/J.COBEHA.2021.04.021

Paolone, G., Angelakos, C. C., Meyer, P. J., Robinson, T. E., & Sarter, M. (2013). Cholinergic Control over Attention in Rats Prone to Attribute Incentive Salience to Reward Cues. *The Journal of Neuroscience*, *33*(19), 8321. https://doi.org/10.1523/JNEUROSCI.0709-13.2013

Pavlov, I. P. (1927). An investigation of the physiological activity of the cerebral cortex. In *Annals of neurosciences*. https://doi.org/10.5214/ans.0972-7531.1017309

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6). https://doi.org/10.1037/0033-295X.87.6.532

Pietrock, C., Ebrahimi, C., Katthagen, T. M., Koch, S. P., Heinz, A., Rothkirch, M., & Schlagenhauf, F. (2019). Pupil dilation as an implicit measure of appetitive Pavlovian learning. *Psychophysiology*, *56*(12), e13463. https://doi.org/10.1111/PSYP.13463

Pittig, A., Heinig, I., Goerigk, S., Richter, J., Hollandt, M., Lueken, U., Pauli, P., Deckert, J., Kircher, T., Straube, B., Neudeck, P., Koelkebeck, K., Dannlowski, U., Arolt, V., Fydrich, T., Fehm, L., Ströhle, A., Totzeck, C., Margraf, J., … Wittchen, H. U. (2023). Change of Threat Expectancy as Mechanism of Exposure-Based Psychotherapy for Anxiety Disorders: Evidence From

8,484 Exposure Exercises of 605 Patients. *Clinical Psychological Science*, *11*(2), 199–217. https://doi.org/10.1177/21677026221101379/SUPPL_FILE/SJ-DOCX-1-CPX-10.1177_21677026221101379.DOCX

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, *306*(5703), 1944–1947. https://doi.org/10.1126/SCIENCE.1102384/SUPPL_FILE/REDISH.SOM.PDF

Robinson, T. E., & Flagel, S. B. (2009). Dissociating the Predictive and Incentive Motivational Properties of Reward-Related Cues Through the Study of Individual Differences. *Biological Psychiatry*, *65*(10), 869–873. https://doi.org/10.1016/j.biopsych.2008.09.006

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. S. (2008). Frontal Cortex Subregions Play Distinct Roles in Choices between Actions and Stimuli. *Journal of Neuroscience*, *28*(51), 13775–13785. https://doi.org/10.1523/JNEUROSCI.3541-08.2008

Saunders, B. T., & Robinson, T. E. (2010). A Cocaine Cue Acts as an Incentive Stimulus in Some but not Others: Implications for Addiction. *Biological Psychiatry*, *67*(8), 730–736. https://doi.org/10.1016/J.BIOPSYCH.2009.11.015

Schad, D. J., Rapp, M. A., Garbusow, M., Nebe, S., Sebold, M., Obst, E., Sommer, C., Deserno, L., Rabovsky, M., Friedel, E., Romanczuk-Seiferth, N., Wittchen, H. U., Zimmermann, U. S., Walter, H., Sterzer, P., Smolka, M. N., Schlagenhauf, F., Heinz, A., Dayan, P., & Huys, Q. J. M. (2020). Dissociating neural learning signals in human sign- and goal-trackers. *Nature Human Behaviour*, *4*(2), 201–214. https://doi.org/10.1038/S41562-019-0765-5

Soltani, A., & Koechlin, E. (2022). Computational models of adaptive behavior and prefrontal cortex. In *Neuropsychopharmacology* (Vol. 47, Issue 1). https://doi.org/10.1038/s41386-021-01123-1

Soltani, A., & Wang, X. J. (2006). A Biophysically Based Neural Model of Matching Law Behavior: Melioration by Stochastic Synapses. *Journal of Neuroscience*, *26*(14), 3731–3744. https://doi.org/10.1523/JNEUROSCI.5159-05.2006

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782–1787. https://doi.org/10.1126/SCIENCE.1094765/SUPPL_FILE/SUGRUE.SOM.REV.PDF

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In *Inhibition and learning*.

Watson, P., Wiers, R. W., Hommel, B., & De Wit, S. (2014). Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite*, *79*, 139–148. https://doi.org/10.1016/J.APPET.2014.04.005

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Woo, J. H., Costa, V. D., Taswell, C. A., Rothenhoefer, K. M., Averbeck, B. B., & Soltani, A. (2024). Contribution of amygdala to dynamic model arbitration under uncertainty. *BioRxiv*, 2024.09.13.612869. https://doi.org/10.1101/2024.09.13.612869

Figure 1. Illustration of the Pavlovian-to-Instrumental Transfer task. (a) Pavlovian learning phase. Participants underwent a series of trials where, following a 5000 ms intertrial interval (ITI), one of three CSs was displayed in the upper display of the slot-machine. Subsequently, the associated outcomes were presented in the lower display. Of the three CSs, two were linked to distinct rewarding outcomes (O_1 and O_2), while the third control CS (CS-) was paired with a non-rewarding outcome. Participants had to learn the CS-outcome associations. (b) Instrumental learning phase. Following an intertrial interval (ITI) ranging from 500 to 1500 ms, two levers appeared on either side of the slot-machine (R_1 and R_2), and participants had the opportunity to obtain two outcomes (O_1 and O_2) by pressing respectively R_1 or R_2. R_1 led to O_1 with a 70% chance, while R_2 led to O_2 with a 30% chance. The optimal decision-making strategy consisted of pressing more R_1 than R_2, because of the insertion of a "baiting rule", such that the longer the participant abstained from choosing a particular option, the greater the probability of receiving a reward for that option. (c) Transfer phase. Following an intertrial interval (ITI) of 500-1500 ms, one of the three CSs was presented alongside two levers on the left and right sides of the slot machine. Participants could collect two different outcomes (O_1 and O_2) by pressing these levers (R_1 and R_2). The outcomes remained visible for 1000 ms. The probability of obtaining the two outcomes remained the same as the instrumental learning phase (R_1 = 70%; R_2 = 30%, with the baiting rule), independently of the cue presented.

Figure 2. Acquisition of Pavlovian and instrumental learning sign- and goal-trackers. (a) Explicit liking scores before and after Pavlovian learning for $CS+_1$ (black dots), $CS+_2$ (black squares), and CS- (black triangles). Vertical bars represent their 95% credible intervals. Overall, data show an increase in liking for $CS+_1$ and $CS+_2$ after Pavlovian learning in both sign- and goal-trackers which is not confirmed for CS-. (b) Delta gaze index (CS+ minus CS–) for sign-trackers (white dot) and goal-trackers (black dot). Vertical bars represent 95% credible intervals. Overall, data show that sign-trackers exhibit a positive delta (greater gaze index toward CS+) whereas goal-trackers show a negative delta (greater gaze index toward CS-). (c) Pupil dilation (baseline corrected in response to $CS+_1$, $CS+_2$, and CS- in sign-trackers (ST, white dots) and goal-trackers (GT, black dots) during Pavlovian learning. Vertical bars represent 95% credible intervals. Overall, an increase to CSs+ compared to CS- is observed only in sign-trackers. (c) Percentage of trial-by-trial responses during instrumental learning, separately for sign-trackers (left) and goal-trackers (right). The y-axis represents the percentage of R_1 (continuous red line) and R_2 (continuous blue line) throughout the task. The shaded regions reflect their 95% confidence interval across participants for each response. Dashed lines represent the expected matching for R_1 (70%) and R_2 (30%). Data are smoothed using a 5-trial moving average. Overall, the figure shows how both groups learned the optimal decision-making strategy throughout the task, increasing the percentage of R_1.

Figure 3. Pavlovian bias in decision-making in sign- and goal-trackers. (a) Posterior model probabilities (PMPc) associated with the four hypotheses (H_1, H_2, H_3, H_4), together with the complement hypothesis (H_c). (b) R_1 (red) and R_2 (blue) percentages in the four CS conditions. Colored dots represent group means, and colored vertical bars represent 95% confidence intervals. Overall, results show maladaptive Pavlovian bias only in sign-trackers (higher posterior model probability for H_1, i.e., reduced percentage of R_1 selectively in the $CS+_2$ condition). (c) Trial-by-trial percentage of R_1 (continuous red line) and R_2 (continuous blue line) throughout the task for each CS. The shaded regions reflect the 95% confidence

interval across participants for each response. Dashed lines represent the expected matching for R_1 (70%) and R_2 (30%). Data are smoothed using a 5-trial moving average. Overall, results show that the difference between sign-trackers and goal-trackers emerges only in the $CS+_2$ condition, particularly as time progresses.

Figure 4. Goodness-of-fit for various computational models, model recovery, and model validation. (a) Bayesian information criteria for different reinforcement learning models. (b) Akaike information criteria for different reinforcement learning models. (c) Model recovery. Plots show the correlations between estimated and actual model parameters, separately for each free parameter of the best model. (d) Model validation. Plots show the percentage of R1 (red) and R2 (blue) responses for simulated and real data, separately for the four transfer phase conditions. Larger dots represent means, and vertical bars represent standard deviations.

Figure 5. Model estimated parameters. (a) Posterior model probabilities associated with the four hypotheses (H_1, H_2, H_3, and H_4), together with the complement hypothesis (H_c). (b) Estimated γ and αPav from the best-fitting model for individual participants (gray dots), separately for sign-trackers and goal-trackers. Black dots represent means, and vertical bars represent 95% confidence intervals. (c) Time course of the estimated CS values for O_1 (red) and O_2 (blue) during the transfer phase, separately for sign- and goal-trackers. Shaded areas represent the 95% confidence intervals.

Table 1. Comparison criteria for different reinforcement learning models

Model	Parameters	AIC	AIC-min(AIC)	BIC	BIC-min(BIC)
i. Pearce-Hall	2	173.98	13.87	179.76	8.10
ii. Rescorla-Wagner	4	162.02	1.91	173.58	1.91
iii. Coupled	4	162.26	2.15	173.81	2.15
iv. Dynamic ω	4	160.87	0.76	172.42	0.76
v. Dynamic ω - γ	4	160.11	0	171.66	0
vi. Dynamic value ω - γ	4	160.21	0.10	171.76	0.10
vii. Dynamic PE ω - γ	4	160.71	0.60	172.26	0.60
viii. Divided by phase	6	162.19	2.08	179.52	7.86
ix. Rescorla-Wagner + -	6	172.69	12.58	190.02	18.36
x. Coupled *cho - unc*	6	165.60	5.49	182.93	11.27
xi. Dynamic ω - ρ	6	164.00	3.89	181.34	9.67

Note: AIC = Akaike information criteria; BIC = Bayesian information criteria

Table 2. Bayesian informative hypothesis

Hypothesis	BFc	PMPa	PMPc
H_1	54.292	0.706	0.698
H_2	1.32	0.017	0.017
H_3	18.108	0.209	0.207
H_4	5.088	0.067	0.067
H_c			0.011

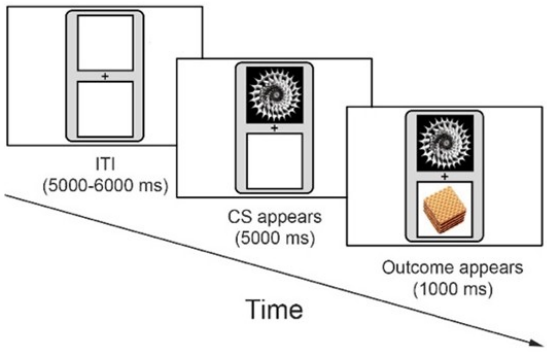
Note: H_c = complement hypothesis; BFc = Bayes Factor versus H_c; PMPa = posterior model probability excluding H_c; PMPc = posterior model probability including H_c

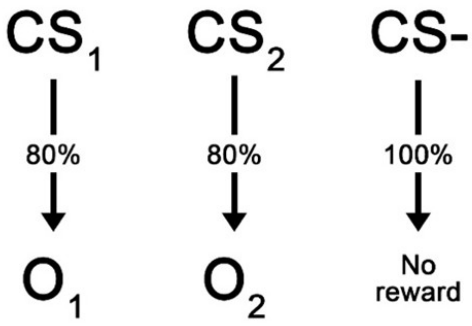
Table 3. Bayesian informative hypothesis

Hypothesis	BFc	PMPa	PMPc
H_1	9.453	0.818	0.756
H_2	0.099	0.009	0.008
H_3	1.203	0.099	0.091
H_4	0.863	0.075	0.069
H_c			0.076

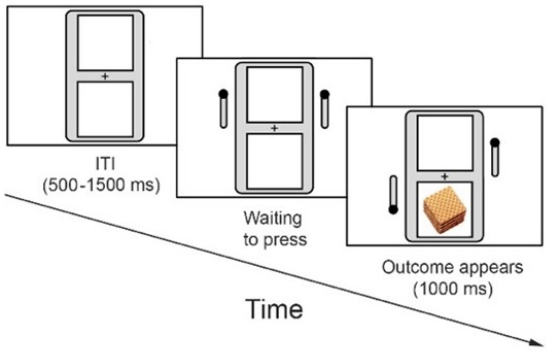
Note: H_c = complement hypothesis; BFc = Bayes Factor versus H_c; PMPa = posterior model probability excluding H_c; PMPc = posterior model probability including H_c

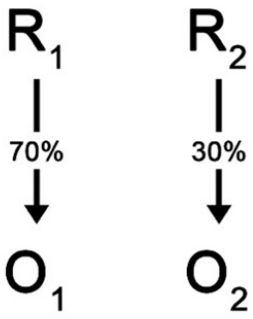
a. Pavlovian learning phase



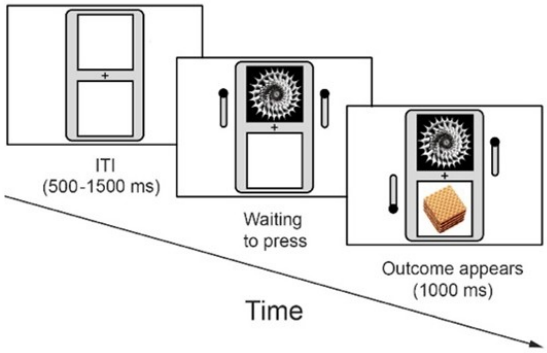


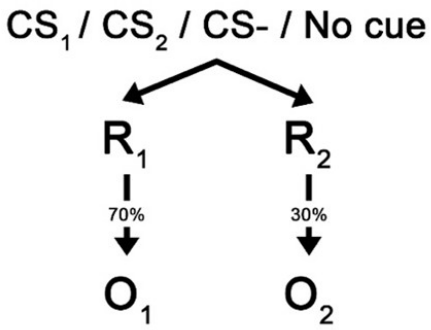
b. Instrumental learning phase

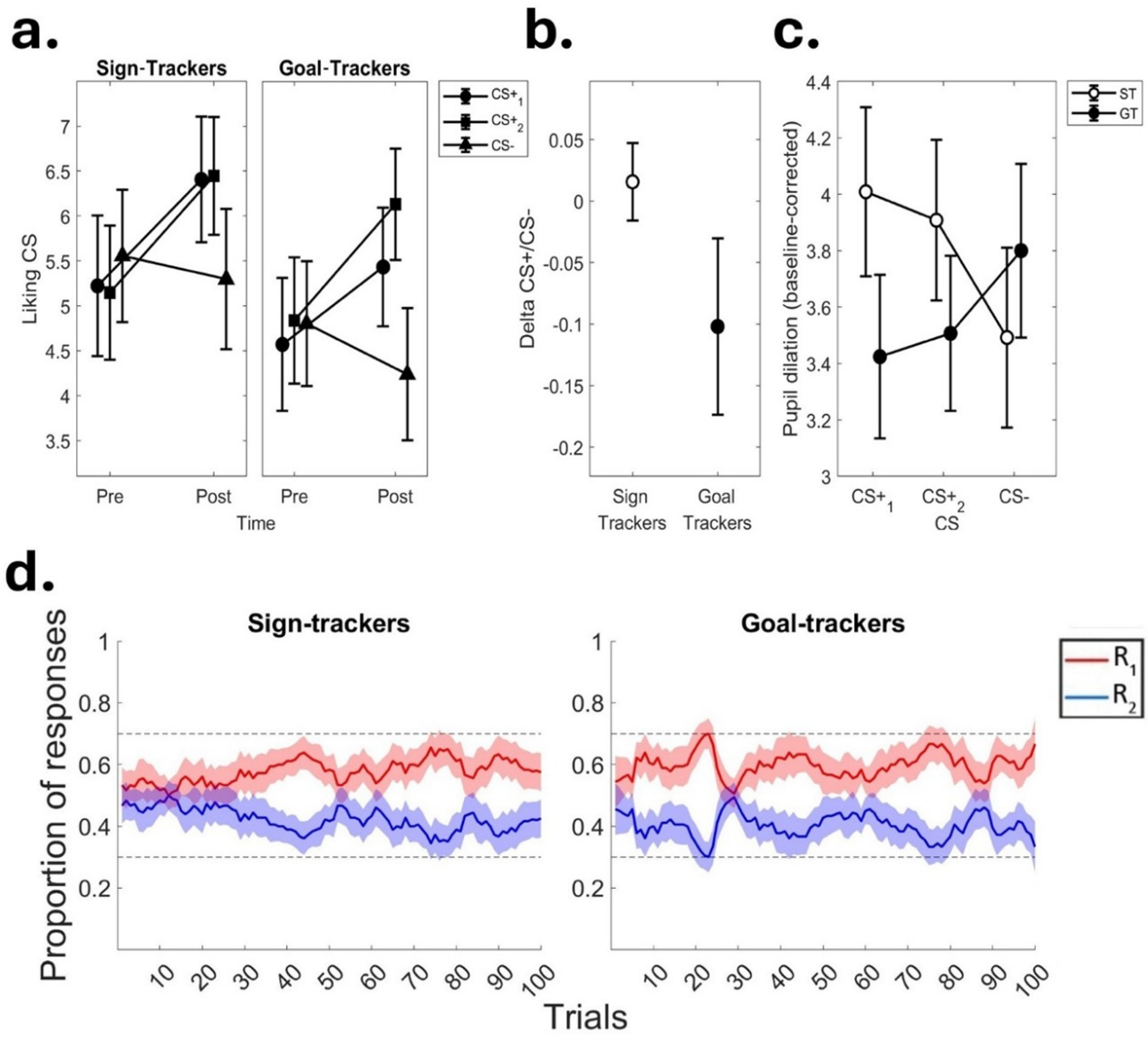


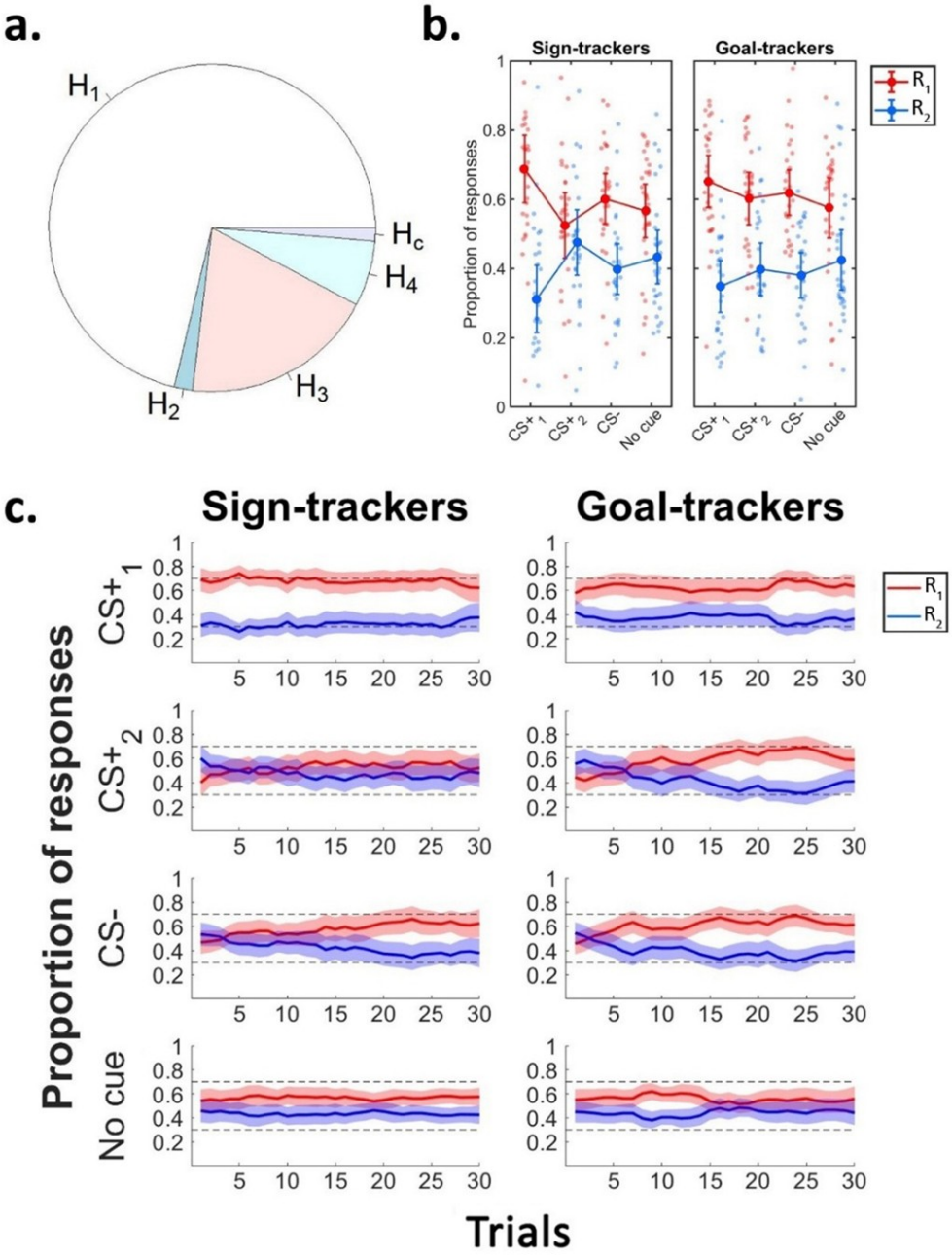


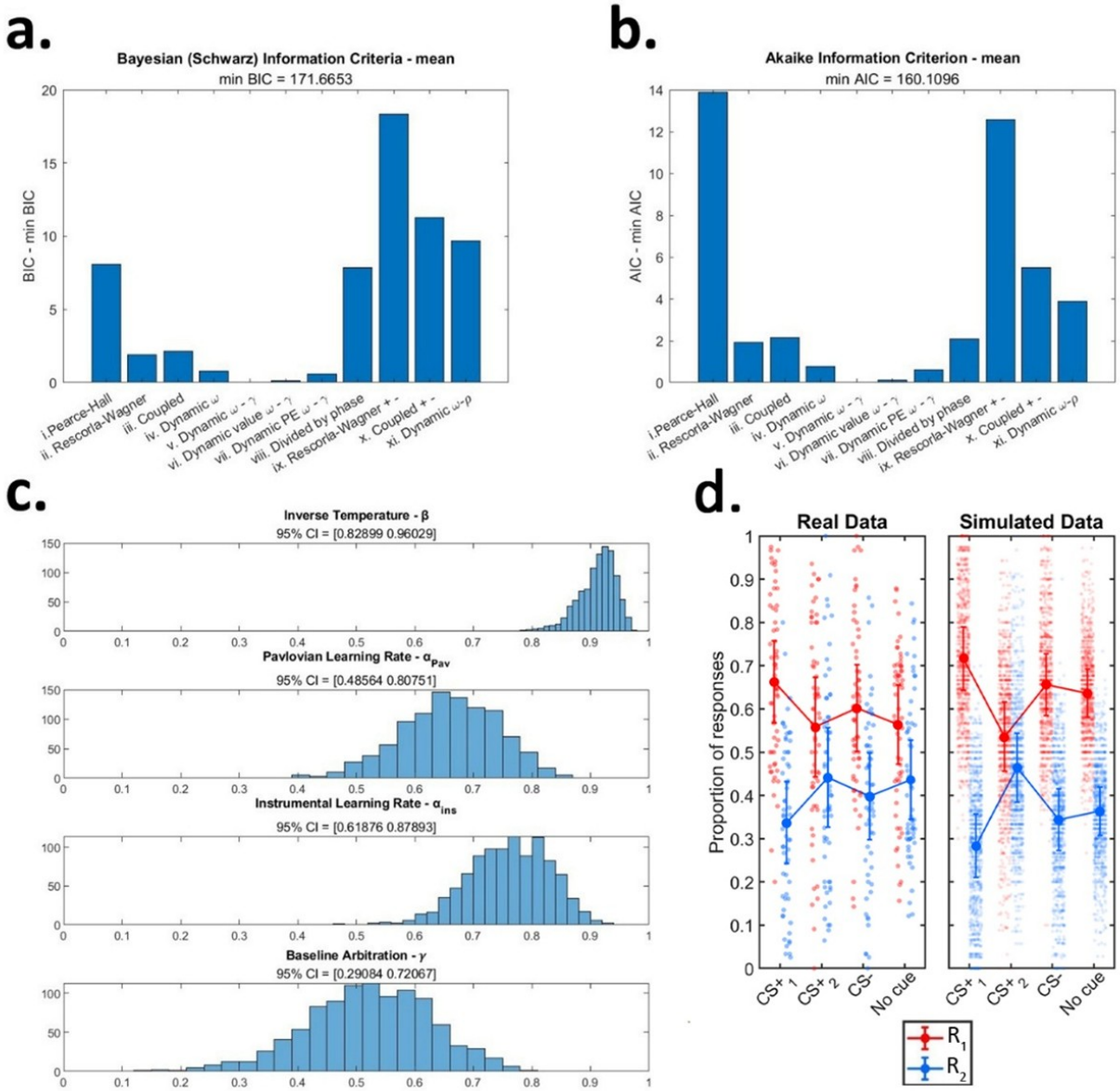
c. Transfer phase

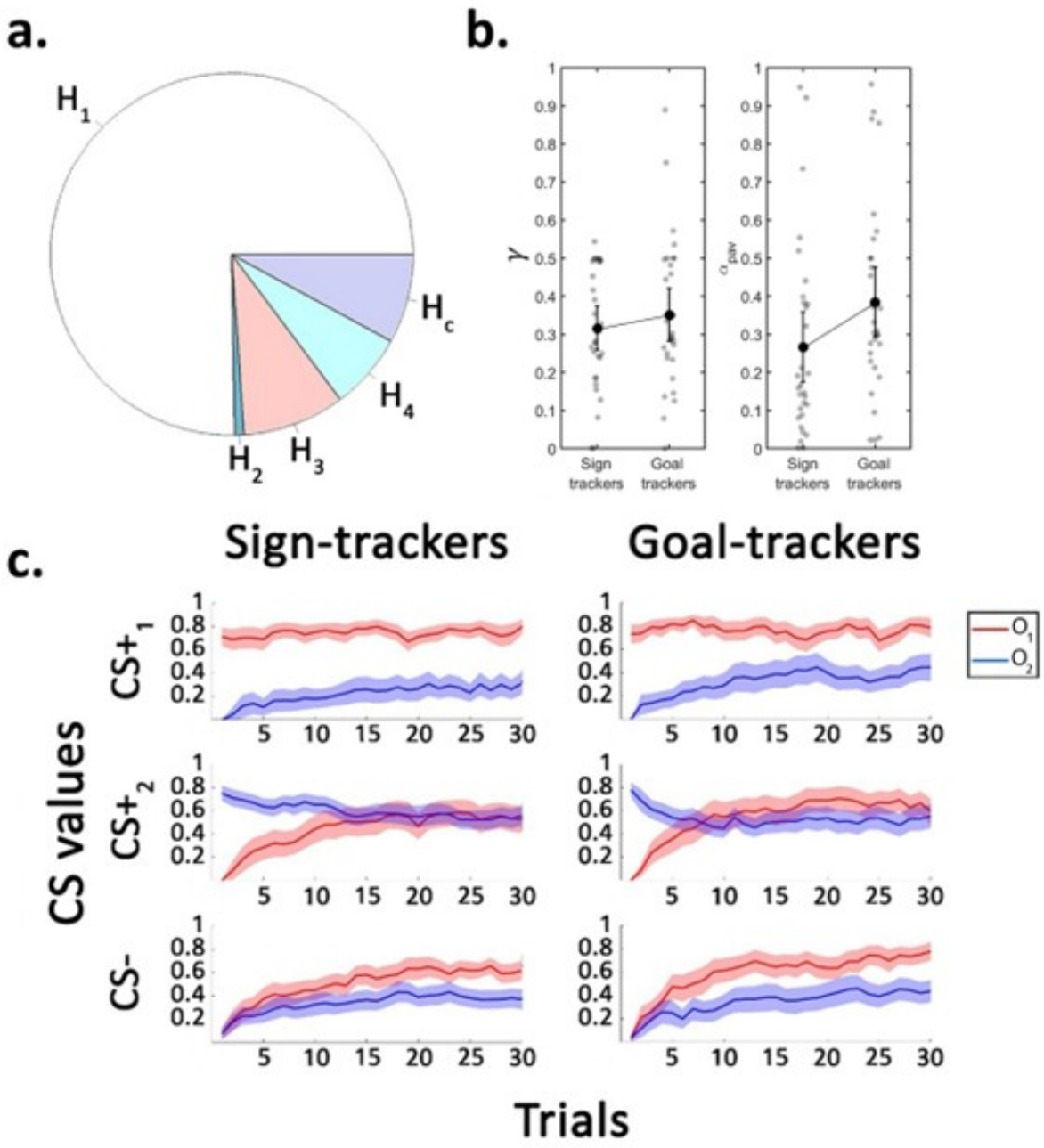





a.
Sign-Trackers
Goal-Trackers
CS+1
CS+2
CS-
Liking CS
Pre
Post
Time
b.
Delta CS+/CS-
Sign Trackers
Goal Trackers
c.
ST
GT
Pupil dilation (baseline-corrected)
CS+1
CS+2
CS-
CS
d.
Sign-trackers
Goal-trackers
R1
R2
Proportion of responses
Trials


a.
H1
Hc
H4
H3
H2
b.
Sign-trackers
Goal-trackers
R1
R2
Proportion of responses
CS+ 1
CS+ 2
CS-
No cue
c.
Sign-trackers
Goal-trackers
Proportion of responses
CS+1
CS+2
CS-
No cue
Trials


a.
Bayesian (Schwarz) Information Criteria - mean
min BIC = 171.6653
BIC - min BIC
i.Pearce-Hall
ii. Rescorla-Wagner
iii. Coupled
iv. Dynamic ω
v. Dynamic ω - ?
vi. Dynamic value ω - ?
vii. Dynamic PE ω - ?
viii. Divided by phase
ix. Rescorla-Wagner + -
x. Coupled + -
xi. Dynamic ω-ρ
b.
Akaike Information Criterion - mean
min AIC = 160.1096
AIC - min AIC
c.
Inverse Temperature - β
95% CI = [0.82899 0.96029]
Pavlovian Learning Rate - α Pav
95% CI = [0.48564 0.80751]
Instrumental Learning Rate - α ins
95% CI = [0.61876 0.87893]
Baseline Arbitration - γ
95% CI = [0.29084 0.72067]
d.
Real Data
Simulated Data
Proportion of responses
CS+ 1
CS+ 2
CS-
No cue
R1
R2


a.
H1
Hc
H4
H2
H3
b.
γ
Sign trackers
Goal trackers
Sign trackers
Goal trackers
c.
Sign-trackers
Goal-trackers
CS values
CS+1
CS+2
CS-
O1
O2
Trials