

UCLA

UCLA Previously Published Works

Title

Delayed reward information is underweighted in reinforcement learning with dispersed feedback

Permalink

<https://escholarship.org/uc/item/7pn3h5xn>

Journal

PLOS Computational Biology, 22(6)

ISSN

1553-734X

Authors

Cotet, Miruna

Poensgen, David

Krajbich, Ian

Publication Date

2026-06-01

DOI

10.1371/journal.pcbi.1014459

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Delayed reward information is underweighted in reinforcement learning with dispersed feedback

Miruna Cotet^{1,2*}, David Poensgen^{3*}, Ian Krajbich^{1,4+}

¹The Ohio State University, Columbus, US

²Complexity Science Hub, Vienna, Austria

³Goethe University Frankfurt, Frankfurt, Germany

⁴University of California, Los Angeles, US

*These authors contributed equally.

+Corresponding author. Email: krajbich@ucla.edu

Abstract

Learning is fundamental to adaptive behavior. In the typical learning task, each action is associated with only one outcome, which could be immediate or delayed. However, actions often have multiple consequences that unfold over time. Here, we used behavioral and eye-tracking experiments to study how people learn when their choices yield both immediate and delayed reward information. Importantly, the rewards themselves were all delivered at the end of the study so there was no reason to weight immediate and delayed reward information differently. Instead, we found that our subjects overweighted immediate reward information. Moreover, this bias increased over the course of the experiment and was still present when learning from others' choices. The gaze data reveal mixed evidence that subjects looked more at immediate vs. delayed feedback, and across subjects, the relative dwell proportion did not predict the behavioral bias. Our results indicate that people prioritize not just immediate rewards, but immediate reward information. Unlike temporal discounting, this form of impatience is a clear mistake and leads to objectively worse outcomes.

Keywords: reinforcement learning, eye-tracking, memory, delay, attention

Author Summary

Decisions often produce several pieces of feedback. Some occur right away, others occur later. Although these signals can be equally useful, people may not treat them that way. In our study, we asked whether humans learn differently from information that arrives immediately versus information that arrives after a delay, even when both are equally valuable. Using a combination of behavioral experiments and eye-tracking, we found that people consistently placed too much emphasis on the immediate feedback. This tendency grew stronger over time and even appeared when participants were learning by watching others' choices. Our findings show that people aren't just drawn to immediate rewards—they are also overly affected by immediate information about those rewards. Unlike classic impatience, where immediate rewards are often genuinely better, immediate information provides no benefit in our setting and overweighting it leads to poor decisions. Our study reveals a systematic and costly quirk in how people learn, and offers an explanation for why people often seem so shortsighted.

Introduction

Learning is fundamental to adaptive behavior [1–3]. To thrive, organisms must learn what things to pursue and what things to avoid. Reinforcement learning (RL) theory [4] is the dominant framework for understanding how organisms do this [5]. In the RL framework, learning involves updating reward expectations based on the difference between experienced rewards and prior expectations. With enough experience, one's reward expectation should converge to the true expected value [6,7].

In the typical RL task, each action is associated with only one reward, which could be immediate or delayed. However, outside the lab, actions often have multiple consequences that unfold over time [8]. In such cases it may be difficult for the organism to learn properly. For instance, credit-assignment is a well-known problem in RL where organisms sometimes misattribute rewards to the wrong actions [9,10]. This problem can occur when there are more actions than outcomes or in complex environments where there is ambiguity about which action(s) led to the reward. What remains unclear is whether people learn appropriately when there is no such ambiguity but there are more outcomes than actions, i.e., when a single action leads to both immediate and delayed rewards.

It is well known that organisms often discount the value of delayed rewards, meaning that they often prefer smaller immediate rewards over larger delayed rewards. This phenomenon is referred to as time discounting or delay discounting and usually thought of as a preference [11–15]. However, it is possible that at least some part of delay discounting is due to inaccurate learning. A difference in the processing of delayed relative to immediate feedback could generate discounting-like behavior that is based on learning frictions rather than preference.

81 There are many reasons why immediate and delayed reward feedback might have
82 different effects on behavior. One reason is limited attention – people may prioritize attending to
83 one kind of feedback over the other. Biased attention can lead to biased choices [16,17].

84 A second reason is agency – people learn better from things they choose compared to
85 things they don't choose [18–21]. Delayed feedback may feel less agentic than immediate
86 feedback. If agency-related boosts in reward prediction error only apply to the most recent
87 choice, then one would expect a preference for options that yield larger immediate rewards.

88 A third reason is memory accessibility – people may find it easier to retrieve immediate
89 feedback than delayed feedback at the time of choice, leading to a stronger influence of
90 immediate vs. delayed feedback on choice. We know that people put more decision weight on
91 information that is presented earlier [22], considered earlier [23,24], or retrieved from memory
92 earlier [25,26]. If immediate feedback comes to mind more quickly than delayed feedback, for
93 instance because of the speed of striatal associations, then this could produce a bias towards
94 overweighting immediate feedback.

95 A fourth reason is an asymmetry in reinforcement learning – people may learn from
96 delayed feedback at a slower rate than from immediate feedback. This possibility arises from
97 research in neuroscience, indicating that immediate and delayed rewards are processed in distinct
98 neural systems. The hippocampus has been shown to be selectively sensitive to delayed
99 feedback, while the ventral striatum has been shown to be selectively sensitive to immediate
100 feedback [27]. Moreover, amnesic patients with damage near the hippocampus are impaired at
101 learning from delayed feedback but not immediate feedback, while patients with striatal
102 dysfunction due to Parkinson's disease are impaired at learning from immediate feedback but not
103 delayed feedback [28,29]. We also know that animals are impaired at learning when there are

longer delays between actions and outcomes [30,31]. Given this evidence that immediate and delayed rewards are processed independently, it seems plausible that these two types of feedback might not be properly integrated into a single, unbiased, total value. While we do not employ brain imaging or patient data in our study, we do explore the implications of two learning systems for behavior.

To study how people integrate dispersed feedback we use a reinforcement learning task in which decision-makers receive feedback about part of the total value immediately after their decision and the second, equally important part, one trial later. Thus, after every decision they need to learn about the option they just chose as well as the option they chose previously. Importantly, the rewards themselves were all delivered at the end of the study session so there was no reason to weight immediate and delayed rewards differently. In some trials, the stimulus with the larger total reward had a smaller immediate reward. Overweighting immediate rewards could lead to errors in these cases. We modeled learning in this task using an extension of a simple RL model [32], embedded in a dynamic choice model [33–36] testing whether our subjects learned from immediate rewards at a higher rate than from delayed rewards. Using eye-tracking, we also tested whether our subjects preferentially gazed at immediate rewards compared to delayed rewards. We correlated the individual learning rate parameters to the proportion of dwell time difference spent on either type of reward to test whether attention could explain any differences in learning rates. We also investigated whether any learning biases were linked to distorted declarative memories for those stimuli. Finally, we also tested whether any learning biases were related to working memory or intertemporal preferences.

To preview our results, we find that people are impaired at learning from delayed feedback, that this immediacy bias is somewhat reflected in aggregate but not individual gaze

measures, and that this behavioral bias may be linked to temporal discounting over much longer timescales. Our subjects put roughly twice as much weight on the immediate feedback as the delayed feedback when making their choices. In the aggregate our subjects tended to look more at the immediate feedback than the delayed feedback, though this tendency did not correlate with the behavioral immediacy bias. Interestingly, the behavioral immediacy bias was still present in a passive learning task where subjects observed others making the first 63 decisions. Finally, subjects who showed a greater immediacy bias in the learning task were, in one experiment, more likely to choose a smaller-sooner than larger-later reward in a standard intertemporal choice task with delays on the order of months. In summary, we find evidence that people do not optimally integrate immediate and delayed feedback and that the bias to discount delayed feedback may be linked to a more general tendency to discount the future.

Results

226 subjects learned the values associated with 6 stimuli, represented by colors in Study 1 (Colors Task; N = 102) and patterns in Study 2 (Patterns Task; N = 124; Fig. 1). An additional 114 subjects completed Studies 3–4 (Patterns Task), reported in the supplementary material. Each stimulus generated a different number of points. Feedback for each choice was not given at once but split into two components: one shown directly after the choice, and one shown after the following trial's choice. Thus, each stimulus had an immediate and a delayed feedback component. Both were equally important in determining the subject's earnings at the end of the study, which were based on the total points earned (SI Methods). Each feedback component had a baseline value (from 2 to 11 points) to which we randomly added up to 4 points in each trial (Fig. 1).

To identify whether subjects would overvalue options with higher immediate than delayed feedback, some stimuli had descending feedback (high immediate and low delayed) while others had ascending feedback (low immediate and high delayed). Each study had two groups of subjects and each group had stimuli that alternated between ascending and descending as the stimuli increased in total reward (Fig. 1).

There were 105 binary choice trials in each study, separated into blocks of 21 trials with every combination of stimuli. We excluded subjects whose accuracy on trials with both options ascending or descending was below 60 percent. We excluded trials in which the two options were identical as there was no (in)correct answer on such trials. We also excluded the first block of trials (21 trials) since subjects had very little information about the stimuli during that block. Finally, we also excluded trials in which subjects did not choose within the time limit (see Methods).

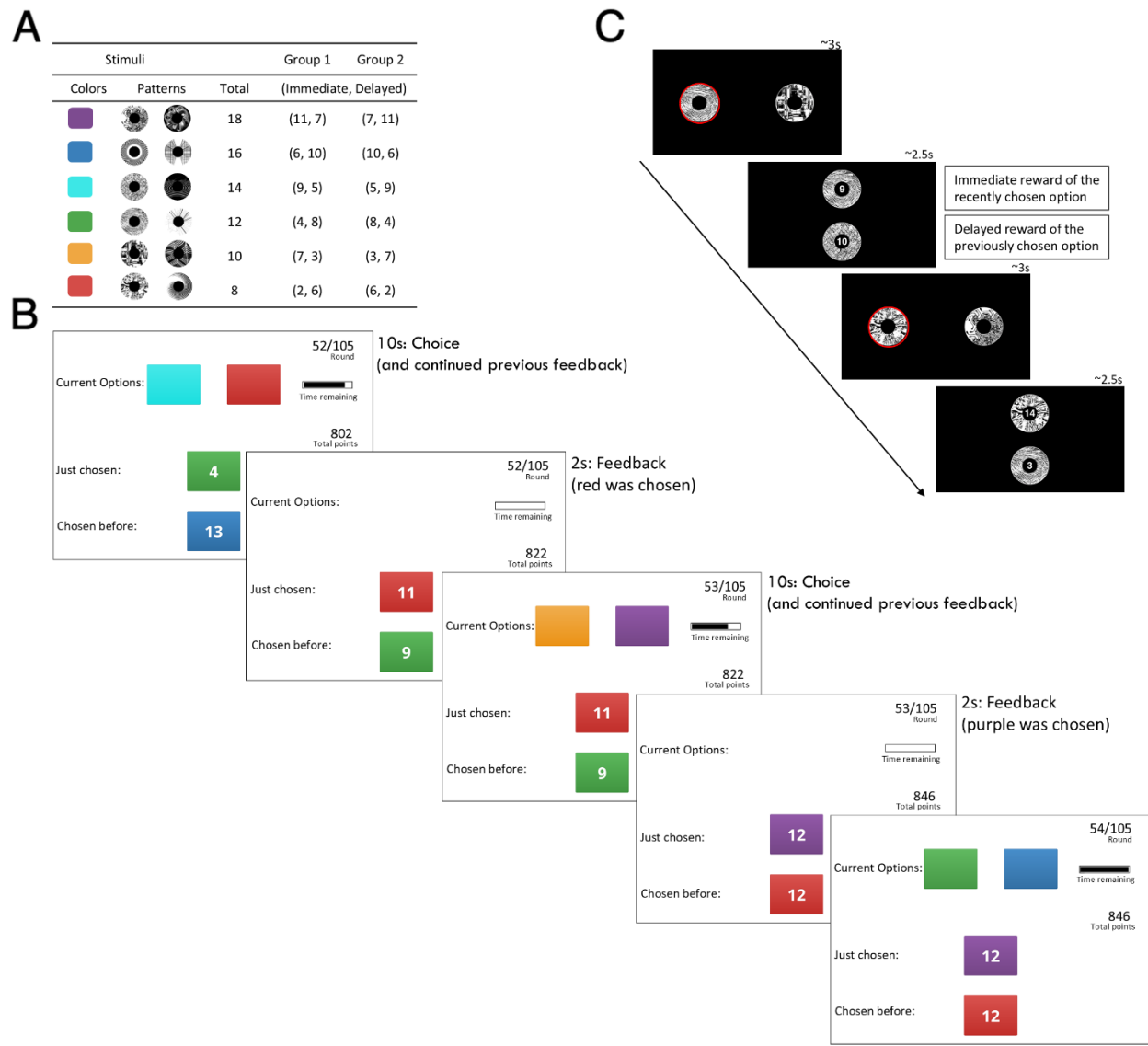


Figure 1. Experimental design. (A) For each subject/block there were six unique stimuli, colors in Study 1 or patterns in Study 2. Each stimulus had a total reward value ranging from 8 to 18, in steps of 2. For each total reward value there was an ascending version and a descending version based on whether the immediate feedback was smaller or larger than the delayed feedback, respectively. The two feedback components were set by dividing the total reward in half, adding 2 for the larger component, and subtracting 2 for the smaller. For each total reward value, Group 1 saw the descending or ascending version and Group 2 saw the opposite. A small random number was added to each underlying feedback component to make learning more difficult. Assignment of colors and patterns to total rewards was randomized by subject. (B) Study 1 timeline. Each trial, subjects chose between two colored rectangles. Below the current options were the immediate

feedback from the previous choice and the delayed feedback from the choice before that. Subjects also saw the trial number, the time within the trial, and the total accumulated points. After each choice, the boxes slid down the screen, feedback was presented, and then new options were shown. (C) Study 2 timeline. Each trial, subjects chose between two patterned discs. After each choice, subjects saw the immediate feedback in the center of their chosen disc. Below (or above) that, they saw the delayed feedback in the center of the chosen disc from the previous trial. Study 2 differed from Study 1 in that the feedback did not continue to be shown during the subsequent choice phase; this change was to facilitate eye tracking.

Behavior

To preview the behavioral results, we find that subjects displayed a learning bias in favor of immediate feedback. Subjects were more likely to choose a descending option than an ascending option, were more likely to make an error when the ascending option was the correct choice, and put more weight on the immediate feedback than the delayed feedback. Moreover, this bias increased over the course of the experiment.

For the following analyses, we have 87 subjects for Study 1 (43 in Group 1, 44 in Group 2) and 90 subjects for Study 2 (47 in Group 1, 43 in Group 2) after exclusions. We also excluded 1 trial in Study 1 and 61 trials in Study 2 due to subjects missing the time limit.

Looking at subjects' choices, we found that they put more weight on immediate feedback compared to delayed feedback. For the same total reward, the option with the higher immediate feedback was more likely to be chosen than its counterpart with the lower immediate feedback (Fig. 2A,B). Incongruent trials (worse option descending, better option ascending) had lower accuracy rates than congruent trials (worse option ascending, better option descending) as evidenced by a paired t-test on the average accuracy rates at the subject level (Study 1: congruent

$M = 0.88$, incongruent $M = 0.53$, $t(86) = -9.28$, $95\%CI = [-0.42, -0.27]$, $p < 10^{-13}$; Study 2: congruent $M = 0.77$, incongruent $M = 0.62$, $t(89) = -4.34$, $95\%CI = [-0.21, -0.08]$, $p < 10^{-4}$).

We confirmed these results using regressions of *Choose Left* on differences in the average immediate, delayed, and total rewards between the left and right options, as well as whether the options were ascending or descending. We used mixed-effects regressions with random intercepts and slopes at the subject level. All continuous variables were z-scored. For each trial we calculated the relevant average feedback seen by the subject up to that point in the experiment. Choosing the left option was more likely when it was descending as opposed to ascending (Study 1: $p < 10^{-13}$; Study 2: $p < .017$; Fig. 2D, Table S1) or when the right option was ascending rather than descending (Study 1: $p < 10^{-14}$; Study 2: $p < 10^{-4}$; Fig. 2D, Table S1), controlling for the difference in total reward. Subjects put larger weights on immediate than delayed feedback. When regressing choice on the immediate and delayed feedback differences, the weight on immediate was higher than on delayed by a factor of 2.4 in Study 1 and 1.4 in Study 2. (Study 1: $\beta_{Immediate} = 2.18$, $95\%CI = [1.97, 2.40]$, $\beta_{Delayed} = 0.90$ [0.71, 1.10]; Study 2: $\beta_{Immediate} = 1.54$ [1.33, 1.75], $\beta_{Delayed} = 1.11$ [0.91, 1.31]; Fig. 2C, Table S2) A Likelihood Ratio Test comparing the immediate and delayed coefficients revealed significant differences (Study 1: $\chi^2(4, N = 87) = 860.74$, $p < 10^{-15}$; Study 2: $\chi^2(4, N = 90) = 350.1$, $p < 10^{-15}$).

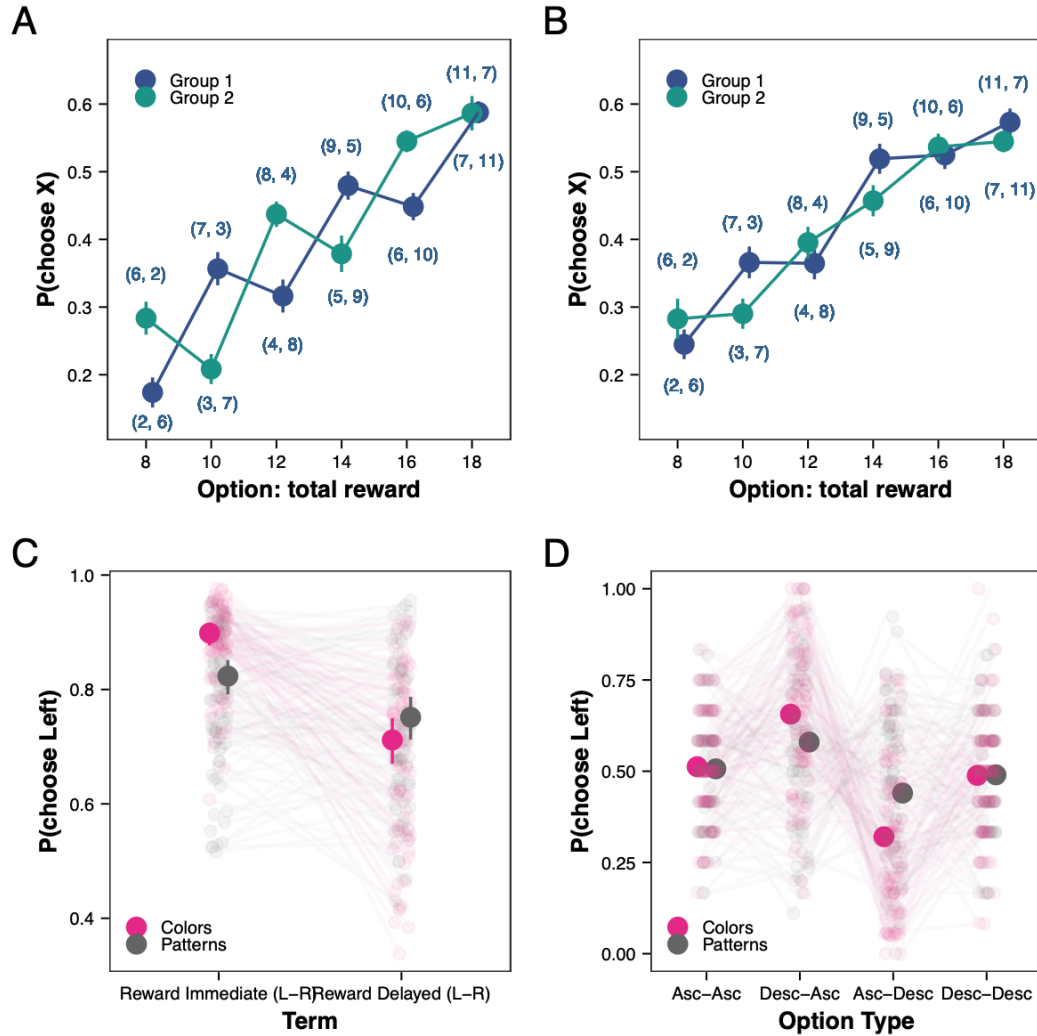


Figure 2. Choice behavior (A,B) Probability of choosing an option as a function of the option's total reward. (A) Study 1 (colors). (B) Study 2 (patterns). (C) Probability of choosing the left option as a function of the difference in the experienced immediate and delayed feedback, based on a mixed-effects logistic regression. Dots represent subject level effects and bars represent standard errors of the fixed effects. (D) Probability of choosing the left option as a function of whether the left and right options were descending or ascending. Dots represent subject level averages and bars represent standard errors across subjects.

The bias to overweight immediate feedback over delayed did not decrease as the experiment progressed; instead, it increased (Fig. 3). Building on the previous regression, we

included interaction effects between trial number and the immediate/delayed feedback. The interaction of trial number and immediate feedback was positive and significant (Study 1: $\beta_{Immediate: Trial} = 0.38$, $95\%CI = [0.28, 0.49]$, $p < 10^{-12}$; Study 2: $\beta_{Immediate: Trial} = 0.34$, $95\%CI = [0.26, 0.43]$, $p < 10^{-15}$; Table S2), while the coefficient for the interaction of trial number and delayed feedback was also significantly positive but smaller (Study 1: $\beta_{Delayed: Trial} = 0.12$, $95\%CI = [0.04, 0.20]$, $p = .003$; Study 2: $\beta_{Delayed: Trial} = 0.19$, $95\%CI = [0.11, 0.26]$, $p < 10^{-6}$; Table S2). A Likelihood Ratio Test comparing the immediate and delayed interaction coefficients revealed that the increase in the delayed coefficient over time was significantly smaller than the increase in the immediate coefficient over time (Study 1: $\chi^2(4, N = 87) = 860.74$, $p < 10^{-15}$; Study 2: $\chi^2(4, N = 90) = 350.1$, $p < 10^{-15}$). This indicates that the immediacy bias increased over the course of the experiment.

The learning bias was also evident in subjects' response times (RT). We know from past work that people make faster choices when there is a larger absolute value difference ($|VD|$) between their options or a larger overall (summed) value (OV) of their options (for appealing options)[37–40]. Since subjects' choices were more influenced by immediate than delayed feedback, we expected to analogously see more influence of immediate than delayed feedback on RT. This was indeed the case. Immediate feedback had a larger effect on RT than delayed for both $|VD|$ and OV. We regressed $\log(RT)$ on $|VD|$ and OV, either in total rewards, or separated into immediate and delayed feedback. Larger total $|VD|$ and larger total OV decreased RT (Study 1: $\beta_{|VD|} = -0.05 [-0.07, -0.04]$, $p < 10^{-13}$, $\beta_{OV} = -0.06 [-0.08, -0.05]$, $p < 10^{-12}$; Study 2: $\beta_{|VD|} = -0.04 [-0.05, -0.03]$, $p < 10^{-9}$, $\beta_{OV} = -0.05 [-0.06, -0.04]$, $p < 10^{-10}$; Figs. 4A,D & S14; Table S3). However, when we separated reward into immediate and delayed components, we found that immediate but not delayed $|VD|$ significantly decreased RT (Study 1: $\beta_{VD_Immediate} = -0.04$

$[-0.06, -0.03]$, $p < 10^{-9}$; $\beta_{VD_Delayed} = -0.002$ $[-0.01, 0.01]$, $p = .786$; Study 2: $\beta_{VD_Immediate} =$
 -0.02 $[-0.03, -0.01]$, $p < 10^{-3}$; $\beta_{VD_Delayed} = -0.01$ $[-0.02, 0.00]$, $p = .056$; Figs. 4B,C & S14,
 Table S4). The difference between immediate and delayed effects was significant (Study 1: $\chi^2(6,$
 $N = 87) = 114.19$, $p < 10^{-15}$; Study 2: $\chi^2(6, N = 90) = 82.67$, $p < 10^{-14}$). We also found that
 immediate OV had a larger effect on RT than delayed OV (Study 1: $\beta_{OV_Immediate} = -0.07$
 $[-0.09, -0.05]$, $p < 10^{-12}$; $\beta_{OV_Delayed} = -0.03$ $[-0.04, -0.01]$, $p < 10^{-4}$; Study 2: $\beta_{OV_Immediate} =$
 -0.04 $[-0.05, -0.03]$, $p < 10^{-8}$; $\beta_{OV_Delayed} = -0.03$ $[-0.04, -0.02]$, $p < 10^{-5}$; Difference: Study 1:
 $\chi^2(6, N = 87) = 95.29$, $p < 10^{-15}$; Study 2: $\chi^2(6, N = 90) = 52.07$, $p < 10^{-8}$; Figs. 4C,F & S14,
 Table S4).

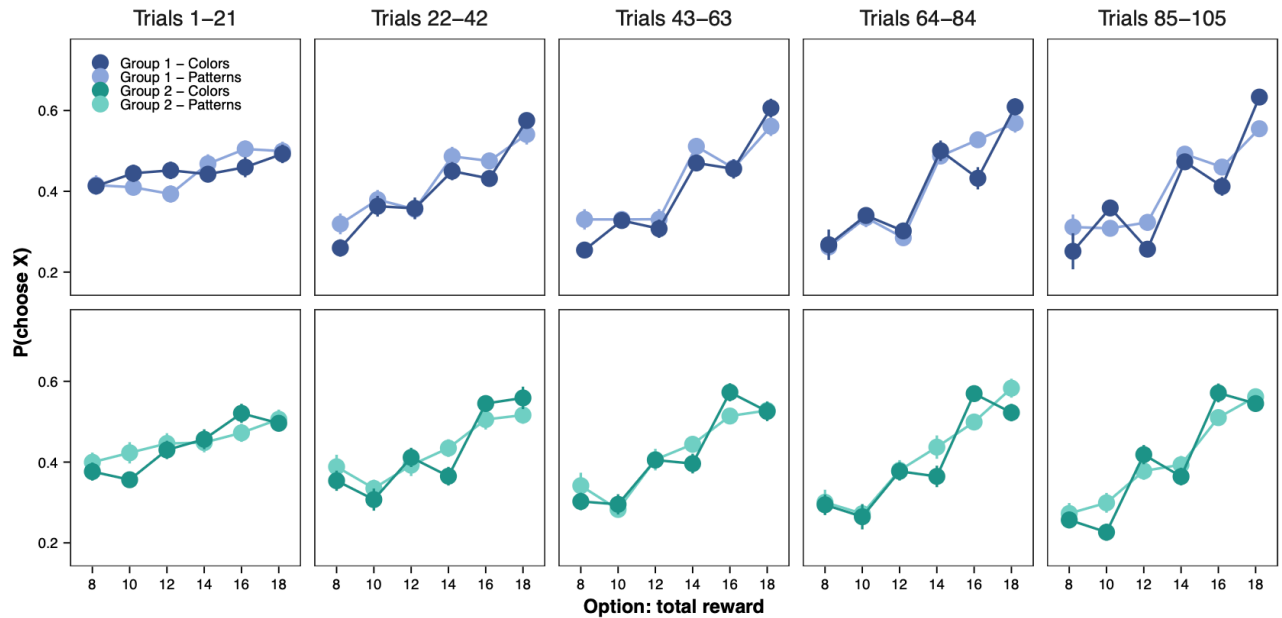


Figure 3. Behavioral bias over time. Probability of choosing a stimulus given it is in the choice
 set as a function of the total reward of the stimulus. The graphs show both the learning of total
 values (average slope) and the immediacy bias (zig-zag pattern) across trials. Though visually
 subtle, statistical analyses reveal that the immediacy bias increases across trials (Table S2). Dots
 represent subject level averages and bars represent standard errors across subjects.

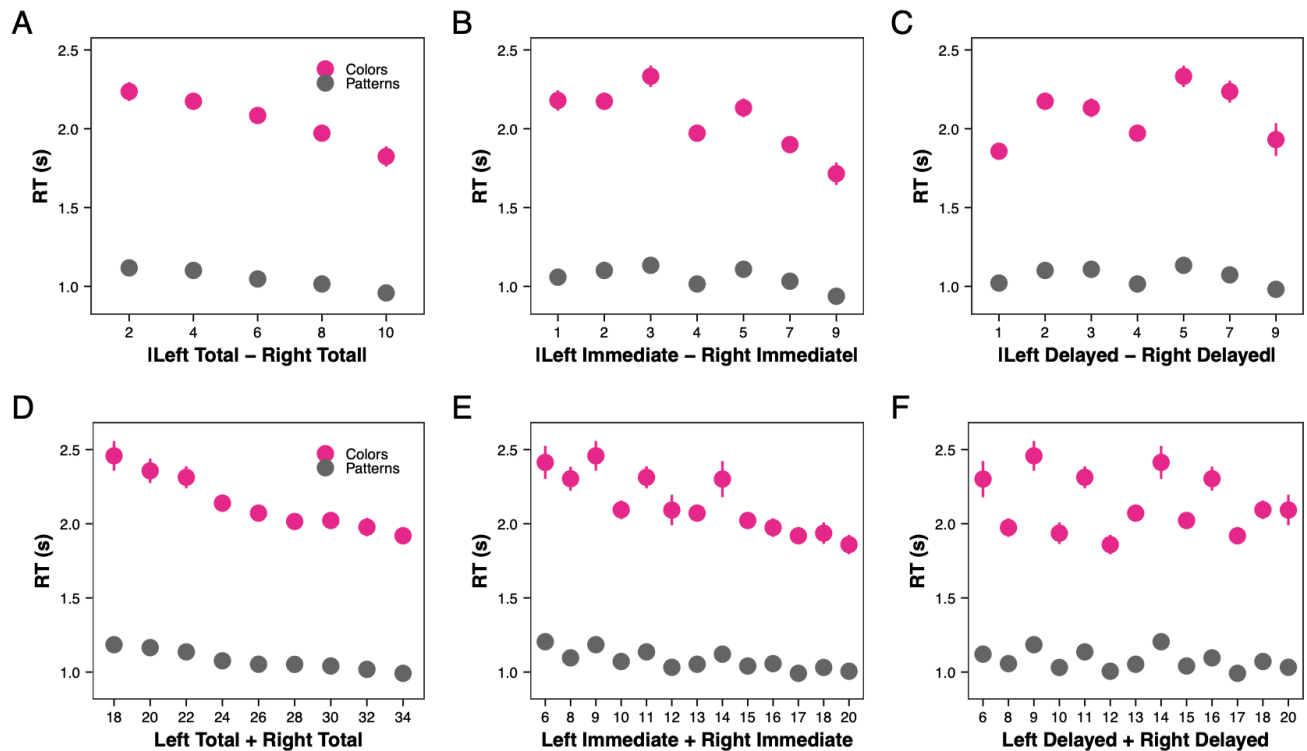


Figure 4. Value difference and overall value effects on response time (RT). (A,B,C) Absolute value difference ($|VD|$) effects on RT for each study. (A) RT decreases with total $|VD|$. (B) RT decreases with immediate $|VD|$. (C) RT does not decrease with delayed $|VD|$. (D,E,F) Overall value (OV) effects on RT for each study. (D) RT decreases with total OV. (E) RT decreases with immediate OV. (F) RT weakly decreases with delayed OV. Dots and bars represent mean and standard errors across subjects.

Computational Model

We sought to capture the relative effect of immediate and delayed rewards using a formal learning model. To do so, we employed an RL model embedded in a drift diffusion model (DDM), but allowed different learning rates for immediate and delayed rewards. We focus on this differential-learning model because our post-task data (described below) indicate an impairment during learning rather than decision-making, but later we also account for possible biases during choice.

We modeled 176 subjects, 87 subjects from Study 1 and 89 subjects from Study 2. One subject was excluded from Study 2's modeling analyses due to too few valid trials.

We fitted a reinforcement learning drift diffusion model (RLDDM) [33–35]. Given that each stimulus (s) was associated with two rewards, one immediate and one delayed (r^I, r^D), the total value for a stimulus was the sum of the predicted values for the immediate and delayed rewards (V). The two values generated two different prediction errors, one for the immediate reward and one for the delayed reward on each trial (k). The learning rates (α^I, α^D) controlled how much the values were updated after each reward. In order to determine if learning occurred at different rates for the immediate and delayed rewards, we also fitted a restricted model in which the learning rate was the same for both rewards ($\alpha^I = \alpha^D = \alpha$).

Consistent with the model-free analyses, we found higher learning rates for immediate rewards than for delayed rewards. This was true for both studies (Study 1: $M_{Immediate} = 0.27$, $M_{Delayed} = 0.14$, $t(86) = 6.97$, $95\%CI = [0.09, 0.16]$, $p < 10^{-9}$; Study 2: $M_{Immediate} = 0.23$, $M_{Delayed} = 0.18$, $t(88) = 2.75$, $95\%CI = [0.01, 0.09]$, $p = .007$; Fig. 5C). According to subject-level WAIC, the model with differential learning provides a better fit for 48% of our subjects while the model with the same learning rate for both types of rewards fit 52% of our subjects better. Thus, about half of our subjects were better described by two different learning rates.

The differential learning model fits the data well in an absolute sense. It reproduces the learning bias observed in the data; the option with the higher immediate reward was more likely to be chosen than the option with the higher delayed reward, for the same total reward level (Fig. S11). The model also shows increased accuracy across trials (Fig. 5A). The model does a good job of predicting accuracy for each type of choice trial (Fig. 5B) and the correlation between subject-level predicted accuracy and observed accuracy was high ($r(174) = .67$, $p < 10^{-15}$; Fig.

S12). The model predicts behavior more accurately on congruent and incongruent trials compared to when both options were ascending or descending (Fig. 5B).

In addition to the differential learning rates, it is also possible that people weight immediate and delayed rewards differently at the time of choice. This could occur either due to unequal attribute weighting [41] or due to discounting delayed feedback, relative to immediate feedback, during learning; these two mechanisms are mathematically equivalent assuming initial values are zero. To account for this possibility, we additionally fit models with different weights on immediate and delayed values, either with or without the different learning rates.

The model with equal learning rates and equal decision weights fit 32% of the participants best. Of the remaining participants, a majority (41%) were best fit by models that included different learning rates (best fits: unequal learning rates (14%), unequal decision weights (28%), or both (27%)). When considering data from all studies (87 for Study 1, 89 for Study 2, 40 for Study 3, 47 for Study 4), again the model with equal learning rates and equal decision weights fit 1/3 (33%) of the participants best. Of the remaining participants, a majority (53%) were best fit by models that included different learning rates (best fits: unequal learning rates (13%), unequal decision weights (31%), or both (22%)) (Table 1).

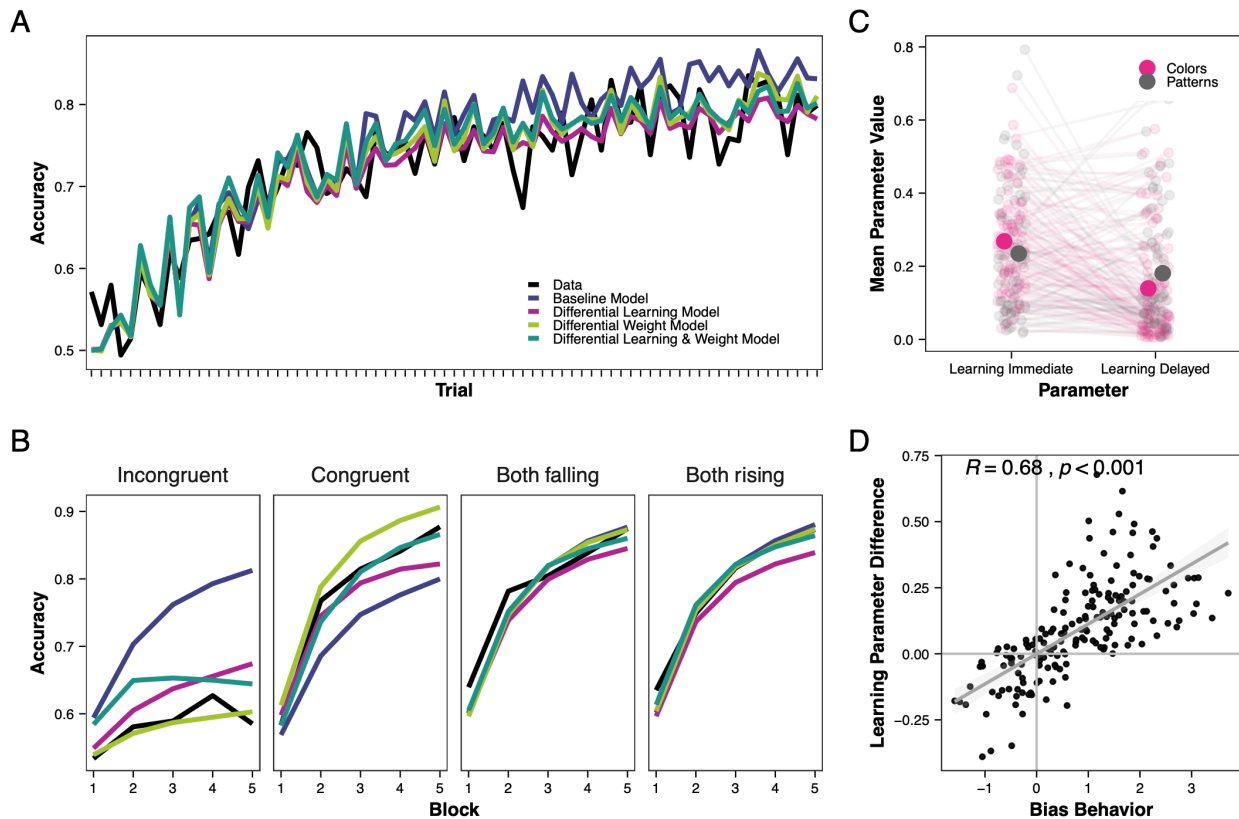


Figure 5. Model parameters and fits. (A,B) Average choice accuracy in the data and in the model simulations using the mean posterior values across trials and subjects. (A) Experiment level. (B) Block level for each type of trial. Incongruent choice sets: worse option descending, better option ascending, Congruent choice sets: worse option ascending, better option descending (C) Means and standard errors of the posteriors for the immediate and delayed learning rates. Dots represent subject-level learning rates for each task. (D) Correlation between behavioral bias and differential learning rate model bias. Behavioral bias is measured as the difference between the immediate reward coefficient and delayed reward coefficient in the regression of choice on rewards. Learning parameter difference is measured as the difference in the learning rates between the immediate and delayed reward.

	N	P	Mean (WAIC - WAIC Best)	Sum WAIC
Differential Weight Model	82	0.31	1.24	61100
Differential Learning and Weight Model	59	0.22	1.58	61198
Differential Learning Model	34	0.13	1.88	61306
Baseline Model	88	0.33	4.79	61715

Table 1. Model comparison results using WAIC. Total sample size from Studies 1-4 was 263 subjects. N and P are the number of subjects and proportion of subjects best fitted by each model. Mean (WAIC – WAIC Best) is the mean increase in WAIC for all the models that were not the best model. Sum WAIC is the sum of all WAIC across subjects.

Eye-tracking

As mentioned in the introduction, attention is one possible explanation for the differential impact of immediate and delayed feedback. To address this, we collected eye-tracking data in Study 2 (and Studies 3–4 in the supplementary material) while subjects saw feedback and made their choices. We sought to test whether subjects allocated more gaze to the immediate or delayed feedback, and whether the relative fraction of dwell time on the two feedback components predicted the learning bias. To preview the results, we observed a tendency to dwell longer on the immediate feedback compared to the delayed feedback, but this gaze bias was not correlated with the behavioral bias.

For these analyses, we used 75 subjects, after excluding 15 subjects who did not pass at least three out of four calibration checks throughout the experiment. We did not exclude any trials.

On the whole, subjects tended to look more at the immediate feedback compared to the delayed one (Fig. 6C). In a mixed-effects regression of relative dwell proportion on immediate vs. delayed, there was a marginal bias in favor of the immediate feedback when controlling for the size and type (ascending vs. descending) of reward ($\beta = 0.05$, 95% $CI = [0.00, 0.11]$, $p = .070$; Bayesian model: posterior mean $\beta = 0.05$, 95% $CrI [-0.005, 0.11]$, $\Pr(\beta > 0) = .96$; Table S6),

and a significant bias when controlling for the predicted values and prediction errors ($\beta = 0.09$, $95\%CI = [0.03, 0.16]$, $p = .007$; Bayesian model: posterior mean $\beta = 0.09$, $95\% CrI [0.02, 0.17]$, $Pr(\beta > 0) = .996$; Table S7). Using subject-level regressions, we found that 21 out of 75 subjects had a significant gaze bias towards the immediate feedback and 14 out of 75 had a significant gaze bias towards the delayed feedback.

Subjects were not more likely to fixate first on the immediate compared to the delayed feedback. In a mixed-effects logistic regression of first fixation location (immediate vs. delayed) there was no significant bias towards the immediate feedback, neither when controlling for the size and type of reward ($\beta = 0.09$, $95\%CI = [-0.12, 0.31]$, $p = .379$; Bayesian model: posterior mean $\beta = 0.10$, $95\% CrI [-0.11, 0.32]$, $Pr(\beta > 0) = .83$; Table S6) nor when controlling for the predicted values and prediction errors ($\beta = 0.21$, $95\%CI = [-0.07, 0.49]$, $p = .138$; Bayesian model: posterior mean $\beta = 0.21$, $95\% CrI [-0.07, 0.49]$, $Pr(\beta > 0) = .93$; Table S7). There was considerable heterogeneity across subjects in first fixation location. Using subject-level logistic regressions, we found that 12 out of 75 subjects had a significant first-fixation bias towards the immediate feedback and 11 out of 75 subjects had a significant first-fixation bias towards the delayed feedback.

Surprisingly, gaze biases did not positively correlate with choice biases. This was true both when using the behavioral bias calculated from the regressions (Dwell Proportion: Spearman's $\rho(73) = -.13$, $p = .27$; Bayesian analysis $\rho = -.12$, $95\% CrI [-0.34, 0.10]$; First Fixation: Spearman's $\rho(73) = -.25$, $p = .028$; Bayesian analysis $\rho = -.24$, $95\% CrI [-0.44, -0.02]$; Fig. S10) and when using the learning rates from the RL model (Dwell Proportion: Spearman's $\rho(73) = -.24$, $p = .04$; Bayesian analysis $\rho = -.19$, $95\% CrI [-0.40, 0.03]$; First Fixation:

Spearman's $\rho(73) = -.23, p = .05$; Bayesian analysis $\rho = -.23, 95\% CrI [-0.43, -0.01]$; Fig. 6A,B). If anything, the correlation was in the opposite direction to what we expected.

However, we did replicate the general finding that options that receive more gaze during the choice phase are more likely to be chosen [16]. A regression of *Choose Left* on the dwell-time difference between left and right options revealed a positive and significant effect controlling for immediate and delayed feedback differences ($\beta = 0.43, 95\%CI = [0.30, 0.56], z = 6.36, p < 10^{-9}$; Table S8).

Studies 3–4 (Study 3 online, Study 4 in lab) yielded slightly different findings. In these studies, the correlations between gaze bias to the immediate feedback and choice bias were positive (rather than negative as in Study 2), but were only significant or marginal in Study 3 when looking at the choice bias from the regressions, and marginal in Study 4 when looking at the learning rates. There was also no significant overall dwell bias towards the immediate feedback in either study, though there was a significant first-fixation bias towards the immediate feedback in Study 4 (Figs. S10, S20–S21). Overall, these additional studies suggest that there may be a link between gaze during the feedback phase and choice bias, but that it is likely a weak relationship.

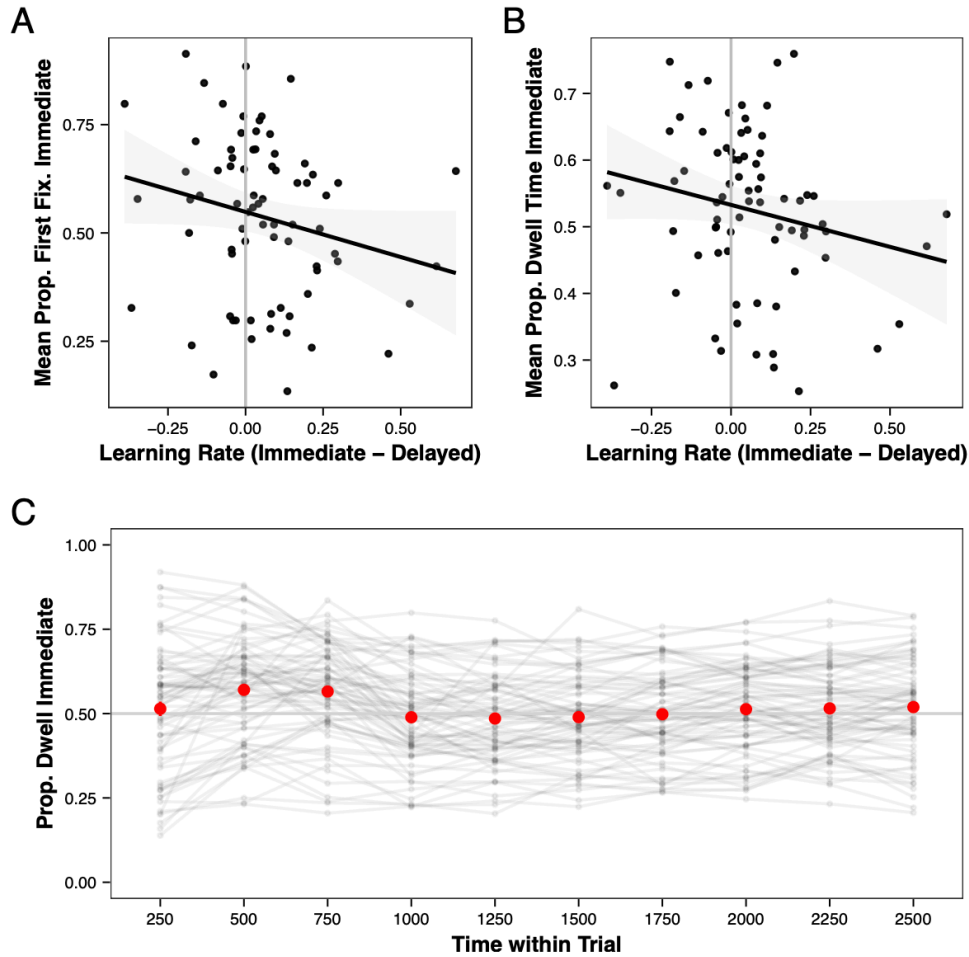


Figure 6. The link between gaze and behavior in Study 2. (A) The correlation between the difference in learning rate (immediate vs. delayed) and the proportion of first fixations to the immediate feedback (Spearman's $\rho(73) = -.23, p = .05$). (B) The correlation between the difference in learning rate (immediate vs. delayed) and the mean dwell proportion to immediate feedback across trials (Spearman's $\rho(73) = -.24, p = .04$). (C) The mean dwell proportion to immediate feedback within a trial for 250 ms time bins across the trial. Black dots represent each subject within a time bin and red bars are standard errors across subjects. (A,B) Dots represent each subject. The gray bands represent 95% CI.

Passive learning task

As described earlier, one possible explanation for the observed immediacy bias is agency. People learn better from things that they choose than things they don't choose, and so it is possible that the immediacy bias is due to an increased sense that immediate vs. delayed feedback is due to participants' choices.

To investigate this possibility, we conducted an additional experiment with passive learning. N=57 participants completed this passive learning experiment, using the color paradigm from Study 1. Participants first observed 3 blocks of choices from a Study 1 participant. The feedback and timing of the feedback was the same as that of the matched partner, except that they could not see the foregone choice option on the choice screen. After observing 63 trials, participants then made 2 blocks (42 trials) of choices on their own.

Participants in the passive learning experiment still exhibited a substantial immediacy bias in their choices. When regressing choice on the immediate and delayed feedback differences, the weight on the immediate feedback was substantially higher than the weight on delayed feedback ($\beta_{Immediate} = 2.62$, 95%CI = [2.11, 3.12], $p < 10^{-15}$, $\beta_{Delayed} = 1.48$ [1.07, 1.88], $p < 10^{-12}$, Fig. 7, Table S2). While the immediacy bias was perhaps smaller than in Study 1 (depending on the analysis, see SI), passive learning clearly does not eliminate the effect (Fig. 7).

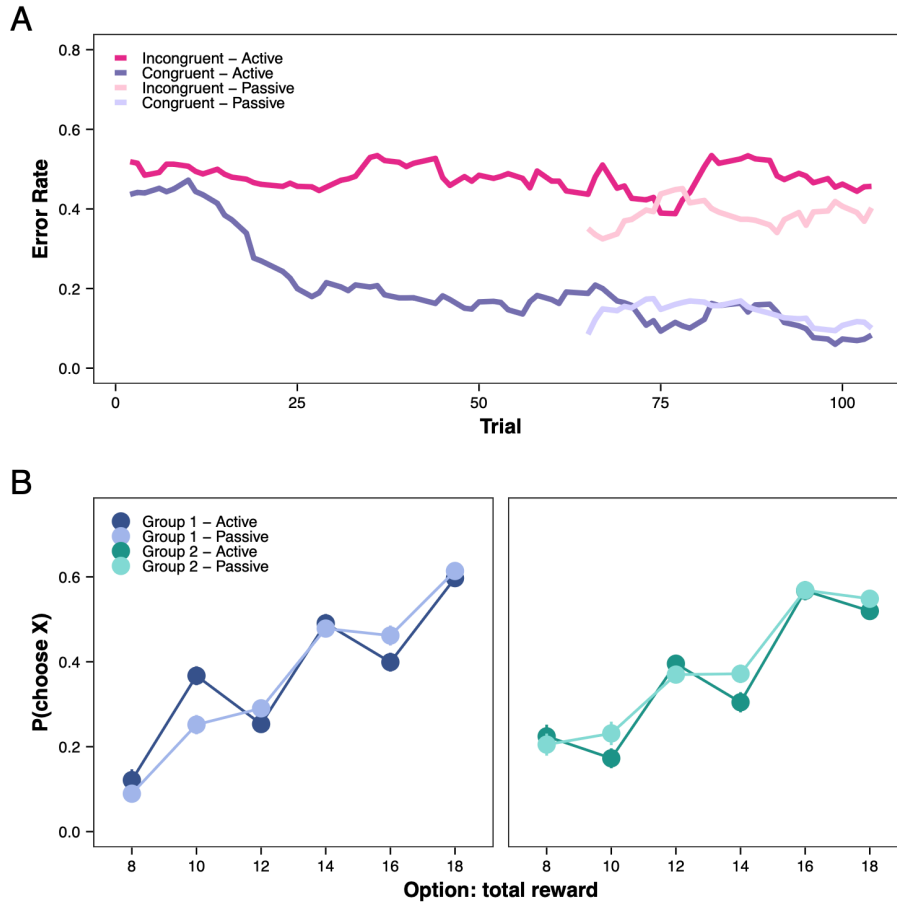


Figure 7. Passive learning experiment (a) Error rates and (b) choice probabilities for the passive learning experiment, compared to Study 1 (active learning matched participants) (a) split into congruent and incongruent trials, and (b) split into Group 1 and Group 2. Darker colors denote the active learning data while lighter colors denote the passive learning data.

Declarative Memory

At the end of Study 2 (and Study 4), we asked subjects to rank the stimuli in terms of total reward, to indicate whether each stimulus was ascending, descending, or flat, and to estimate the average total reward for each stimulus. The results from these extra measures indicate that the behavioral immediacy bias is at least partly due to inaccurate learning, with biases in memory generally biased in the same direction as subjects' choices during the main task.

Rankings

Subjects' rankings were more aligned with the immediate rewards than with the delayed rewards. Using the same ranking responses, we separately computed Kendall tau distance relative to the true immediate-reward ordering and the true delayed-reward ordering. The average ranking error relative to the immediate rewards was 3.23 (Study 4: $SD = 2.1$) and 4.46 (Study 2: $SD = 2.57$), and relative to the delayed rewards was 5.76 (Study 4: $SD = 2.3$) and 5.89 (Study 2: 2.26). These differences are significant (Study 4: $M = -2.52$, $t(45) = -2.52$, $p < .001$; Study 2: $M = -1.43$, $t(107) = -4.02$, $p < .001$). This indicates that subjects encoded the rank order of stimuli primarily through the immediate reward components.

To examine how ranking errors related to choices, we used these same derived measures to separately predict accuracy and bias. A subject who prioritizes immediate rewards should have better alignment of their rankings with the immediate rewards and should behaviorally display more immediacy bias, worse accuracy on incongruent trials, and better accuracy on congruent trials. This is indeed what we observed. Subjects with rankings that were more aligned with the immediate rewards had higher immediacy bias in both studies (linear regression of immediacy bias on ranking error relative to immediate rewards: Study 2: $\beta = -0.24$, 95%CI = $[-0.35, -0.14]$, $p < 10^{-4}$; Study 4: $\beta = -0.25$, 95%CI = $[-0.46, -0.05]$, $p = .02$; Table S10). Subjects with rankings that were more aligned with the immediate rewards also had worse choice accuracy on incongruent trials (linear regression of incongruent error rate on ranking error relative to immediate rewards: Study 2: $\beta = -0.04$, 95%CI = $[-0.07, -0.02]$, $p < .001$; Study 4: $\beta = -0.08$, 95%CI = $[-0.12, -0.04]$, $p < .001$; Table S9), and better accuracy on congruent choice sets (linear regression of congruent error rate on ranking error relative to immediate rewards: Study 2: $\beta =$

0.04, 95%CI = [0.02, 0.06], $p = .001$; Study 4: $\beta = 0.05$, 95%CI = [0.02, 0.08], $p < .001$; Table S9).

Alignment with the delayed-reward rankings had somewhat the opposite pattern. Subjects with rankings that were more aligned with the delayed rewards had lower immediacy bias in one study, and higher immediacy bias in the other (linear regression of immediacy bias on ranking error relative to delayed rewards: Study 2: $\beta = 0.10$, 95%CI = [0.003, 0.201], $p = .04$; Study 4: $\beta = -0.25$, 95%CI = [-0.43, -0.07], $p = .008$; Table S10). Subjects with rankings that were more aligned with the delayed rewards had better choice accuracy on incongruent trials in one study, but not the other (linear regression of incongruent error rate on ranking error relative to delayed rewards: Study 2: $\beta = 0.05$, 95%CI = [0.02, 0.07], $p < 10^{-4}$; Study 4: $\beta = 0.02$, 95%CI = [-0.02, 0.05], $p = .33$; Table S9). There was no significant relation between rankings that were more aligned with the delayed rewards and choice accuracy on congruent trials (Table S9).

In total, these results indicate a relation between subjects' immediacy bias during the choice task and the alignment of their post-experiment rankings with the immediate rewards.

Reward comparisons

In the reward comparison survey, subjects indicated which feedback component (immediate or delayed) was larger for each stimulus. The prediction for this measure was less obvious. On the one hand, better memory for the rewards should reduce any choice bias. On the other hand, better memory for the distinction between immediate and delayed rewards might indicate more disaggregated processing, which combined with favoring the immediate component, would lead to more immediacy bias. Our findings are more aligned with the latter story, with better memory for this comparison associated with a higher immediacy bias (linear

regression of immediacy bias on comparison error: Study 2: $\beta = -0.27$, 95%CI = [-0.46, -0.09], $p = .005$; Study 4: $\beta = -0.26$, 95%CI = [-0.55, 0.03], $p = .008$; Table S10), higher accuracy on congruent choice sets (linear regression of congruent error rate on comparison error: Study 2: $\beta = 0.05$, 95%CI = [0.01, 0.09], $p = .03$; Study 4: $\beta = 0.04$, 95%CI = [-0.004, 0.075], $p = .078$; Table S9), and insignificantly lower accuracy on incongruent choice sets (linear regression of incongruent error rate on comparison error: Study 2: $\beta = -0.02$, 95%CI = [-0.07, 0.02], $p = .22$; Study 4: $\beta = -0.05$, 95%CI = [-0.11, 0.04], $p = .068$; Table S9). Accurate comparative memory may therefore reflect immediate-reward-dominant encoding rather than correcting for it.

Point estimates

In Study 4, subjects separately estimated the average immediate and delayed reward for each stimulus. A subject who prioritizes immediate rewards should have better estimates of the immediate rewards than the delayed rewards. Indeed, subjects made larger absolute errors estimating delayed than immediate rewards (immediate: $M = 1.98$, $SD = 0.51$; delayed: $M = 2.68$, $SD = 0.73$; $t(45) = -4.97$, $p < 10^{-4}$), consistent with poorer encoding of delayed reward information. Accuracy in estimating immediate and delayed rewards was associated with higher accuracy on incongruent sets (Immediate: $\beta = 0.18$, 95%CI = [0.06, 0.31], $p = .005$; Delayed: $\beta = 0.12$, 95%CI = [0.03, 0.22], $p = .01$; Table S9), but not significantly associated with immediacy bias (Immediate: $\beta = -0.37$, 95%CI = [-0.99, 0.26], $p = .25$; Delayed: $\beta = 0.04$, 95%CI = [-0.45, 0.53], $p = .86$; Table S10). When we replaced absolute with signed errors, effects were largely non-significant and did not yield consistent or interpretable patterns, indicating that it is the magnitude rather than the direction of memory error that drives these associations. Thus, subjects

indeed displayed better memory for immediate than delayed rewards, though memory for the reward components did not correlate with the immediacy bias.

Summary

Across our memory measures, a fairly consistent picture emerges: our subjects displayed better memory for immediate than delayed rewards, and to some extent, this correlated with immediacy bias in the choice task. Also, better memory for the distinction between immediate and delayed rewards correlated with immediacy bias. This pattern suggests that the immediacy bias is not a product of inattention or general memory failure, but reflects asymmetrically better encoding of immediate reward information.

Working Memory

At the end of Study 1, subjects also completed a working-memory task – an n-back task using either n=2 or n=3. We regressed the accuracy and d-prime measures for both n-back tasks on the immediacy bias. For the 2-back task, accuracy was marginally negatively related to the immediacy bias based on the regression approach, and significantly so based on the RL model (Regression-based: $\beta = -2.85 [-5.98, 0.28]$, $p = .074$, RL-based: $\beta = -0.55 [-1.02, -0.09]$, $p = .019$; Table S11). The d-prime measure was also negatively related to the immediacy bias using both measures (Regression-based: $\beta = -0.31 [-0.60, -0.03]$, $p = .03$, RL-based: $\beta = -0.05 [-0.10, -0.01]$, $p = .012$; Table S11). However, the association was not present for the 3-back task (Regression-based: $\beta = -0.12 [-0.54, 0.30]$, $p = .556$, RL-based: $\beta = -0.02 [-0.09, 0.05]$, $p = .519$; Table S11).

In Study 4, subjects instead completed a visual working memory task, namely a change localization task. There was no relationship between subjects' performance on that task and their immediacy bias (see supplementary material).

Discounting Preferences

At the end of Study 1 (and Study 4), subjects completed a hypothetical survey to assess their intertemporal preferences. We then regressed the intertemporal indifference point for all three time-scales on the immediacy bias. The only significant association was for the decisions between today and one month. The stronger the behavioral bias, the higher the impatience, significantly for the regression-based analysis ($\beta = 2.04$, 95%CI = [0.15,3.92], $p = .034$; Table S12) and marginally for the RL-based analysis ($\beta = 11.04$ [-1.69,23.78], $p = .088$; Table S12). However, this was marginal or not significant for the other time-scales: today and 6 months (Regression-based: $\beta = 1.86$ [-0.33,4.05], $p = .096$, RL-based: $\beta = 8.03$ [-6.79,22.84], $p = .284$; Table S12) and 1 month and 6 months (Regression-based: $\beta = 1.53$ [-0.54,3.59], $p = .146$, RL-based: $\beta = 3.63$ [-10.36,17.632], $p = .607$; Table S12).

None of these relationships were significant in Study 4 (Table S13). This casts doubt on the relationship between intertemporal preferences and immediacy bias in our learning task.

Discussion

In this article we found that people sub-optimally overweight immediate reward feedback relative to delayed reward feedback. In our task, subjects merely had to add together two (small) numbers to determine the total reward from their choice. And yet, subjects put between 1.4 and 2.4 times as much weight on the immediate feedback compared to the delayed feedback, roughly

equivalent to counting the immediate feedback twice. This resulted in 15 to 35% (Study 1: 35%, Study 2: 15%, Study 3: 17%, Study 4: 25%) increases in errors when the option with the larger immediate feedback was the wrong choice. Surprisingly, this immediacy bias only grew stronger as the experiment progressed. At the same time, we also observed an attentional bias towards immediate feedback – people spend about 53% of the time looking at the immediate feedback (Studies 2 and 3: 52%, Study 4: 54%). However, this gaze bias generally does not correlate with the choice bias. We also found mixed evidence that this immediacy bias may be linked to impatience and working memory deficits.

Our analyses, which focused on both choice and RT, consistently revealed a stronger influence of immediate feedback on subjects' choices. This behavioral bias favoring immediate feedback was further evident in RT, with immediate feedback exerting a larger effect on RT via both value difference and overall value. While RT significantly decreased in both the sum of and the difference between the immediate feedback components of the two options, it showed reduced or no such relationship with the delayed feedback, respectively. The relationship between value-difference, overall value, and RT has been well documented [16,17,34,37–40,42–48] and so its absence/reduction for delayed feedback lends further credence to the idea that people prioritize the immediate feedback.

Overall, we find substantial support for an asymmetry in reinforcement learning whereby people learn more slowly about delayed feedback than immediate feedback. We find mixed support for other possible mechanisms – limited attention, agency, and memory accessibility. Our eye-tracking data indicate that limited attention is unlikely to be the whole story. We find only weak evidence that people fixate more on immediate vs. delayed feedback. Our passive-learning data indicate that agency is unlikely to be the whole story. We still find a substantial

immediacy bias when participants were yoked through others' choices, though the bias did decrease somewhat. Finally, our memory tests at the end of the experiments indicate that unequal weighting of immediate vs. delayed feedback is unlikely to be the whole story. Our participants were worse at recalling delayed vs. immediate feedback, and for a majority of them, their behavior was better fit by a model that incorporated different learning rates. Thus, while attention, agency, and memory accessibility might each marginally contribute to an immediacy bias, we conclude that the bulk of the bias is due to asymmetric learning.

We used an RL model embedded in a DDM to better understand and quantify the prioritization of immediate feedback. We incorporated distinct learning rates for immediate and delayed feedback, which provided a better fit for approximately half of the subjects. Immediate learning rates were significantly higher than delayed learning rates at the group level. Although the two-learning-rate model provided generally good fits to the behavior, it isn't an entirely satisfying explanation for the immediacy bias because while the model can produce growing immediacy bias over short time horizons, it cannot reproduce the continued growth we observed in the data. To confirm this, we generated simulated datasets from the model using fitted subject parameters with low, medium, and high learning rates (all with an immediate:delayed ratio of 2:1), and applied the same mixed-effects logistic regression analysis used on the real data. Across all conditions, the predicted interaction difference ($\beta_{Immediate: Trial} - \beta_{Delayed: Trial}$) was centered near zero, while empirical values from Study 1 (0.26) and Study 2 (0.15) fell outside these distributions (all $p < 0.02$). Moreover, the model under-predicts accuracy when both options are ascending or descending (particularly ascending). Future work will need to investigate other mechanisms by which this immediacy bias might occur.

Consistent with a bias in learning, during the feedback phase we observed some evidence for a greater focus on immediate feedback. In Study 2 subjects spent a greater proportion of time dwelling on immediate vs. delayed feedback, while in Study 4 subjects were more likely to look at the immediate feedback first. However, these attentional biases showed inconsistent correlations with the behavioral bias. In Study 2 the correlations were insignificantly negative, while in Studies 3-4 the correlations were positive but only sometimes significant. The eye-tracking results from Studies 2-3 do need to be interpreted with some caution as the data were collected using online webcam-based eye tracking. Collecting eye-tracking data online has many advantages but it does sacrifice some precision relative to in-lab eye-tracking systems [41,49,50]. Thus, we put more weight on our Study 4 results, which revealed a marginally positive correlation between dwell proportion on the immediate feedback and the learning rate advantage for the immediate feedback. Overall, gaze may play a small role in the immediacy bias but it doesn't seem to fully explain the behavior.

Also consistent with a bias in learning, we found that subjects with a stronger immediacy bias in their choices were also better on the end-of-study memory test where they were asked to report whether each stimulus was ascending, descending, or flat. Moreover, their memory for the rankings or points for the stimuli were, if anything, positively correlated with the immediacy bias. These results indicate that the immediacy bias we observed was not simply due to a lack of attentiveness. Subjects displaying stronger immediacy bias were better at recalling the immediate feedback but not the delayed feedback, suggesting that the immediacy bias may be due to overweighting immediate feedback rather than underweighting delayed feedback.

The act of choosing the alternative that has the higher immediate reward instead of considering how current choices affect future rewards has been termed melioration [51–53].

633 Melioration at the expense of maximization has been demonstrated in many animal experiments
634 [54,55] as well as with human subjects [56]. Previous research has used the Harvard game to
635 show the tendency of melioration over maximization [57–62]. However, the task puts subjects in
636 a highly complex learning environment in which any deviation from the optimal choice appears
637 impatient. Even Bayesian agents with perfect memory often need thousands of trials to avoid
638 melioration and arrive at the optimal solution. The reason lies in the opacity and complexity of
639 that task. Our task explicitly addresses this criticism by being much simpler and fully transparent
640 and allows us to cleanly attribute the observed bias to overweighting of immediate feedback
641 [63].

642 Our results are consistent with research in neuroscience that has argued for different
643 learning systems based on immediate and delayed feedback. The dopamine RL response to
644 delayed rewards is similar to rewards that are entirely unpredicted [64,65]. This indicates that the
645 midbrain-striatal system is not entirely able to learn from delayed rewards [66]. Therefore,
646 learning of delayed rewards might be supported by a different system [27]. There is also
647 evidence for distinct involvement of the striatum and hippocampus for different types of
648 feedback. This research argues that learning from immediate rewards is supported by an implicit
649 memory system associated with the ventral striatum (VS), while learning from delayed rewards
650 is supported by an explicit memory system associated with the hippocampus [27,28,66–68]. Both
651 neuroimaging and patient data support the dissociation between these two learning systems –
652 Parkinson’s patients (with VS dysfunction) are impaired at learning from immediate but not
653 delayed feedback, while amnesic patients (with hippocampal dysfunction) are impaired at
654 learning from delayed but not immediate feedback [27–29]. There is some debate about whether
655 such evidence conclusively supports multiple systems [69]. It is also unclear whether this multi-

system account can explain our results. Unlike these studies, we re-displayed the chosen stimuli when showing both immediate and delayed feedback. So, while the second feedback was delayed relative to the action, it was immediate relative to the stimulus. In any case, future work will need to examine the neural mechanisms underlying the immediacy bias presented here.

Another possibility is that delayed feedback is processed through a Pavlovian rather than instrumental learning system. In the literature, there is a distinction between Pavlovian learning, which depends on stimulus-outcome relations, and instrumental learning, which depends on stimulus-action-outcome relations [70]. In our design, delayed feedback is always the consequence of the participant's action, making it inappropriate to characterize it as purely Pavlovian. Nevertheless, it is possible that delayed feedback is processed by a Pavlovian learning system that takes over when feedback is temporally distant from the action. Our passive control condition partially addresses this possibility: because participants observed outcomes without acting, any immediacy bias that persists across both active and passive conditions cannot be entirely attributed to differences between instrumental and non-instrumental learning. That said, we acknowledge that fully dissociating the effect of delay from the effect of agency is inherently difficult, since delayed feedback is by definition tied to an action. We did consider additional experiments where feedback is presented for non-chosen stimuli. However, we faced two major challenges. First, presenting feedback for non-chosen stimuli would eliminate subjects' incentives to make good choices. Solving this problem would lead to major departures from our design. Second, and more importantly, we determined that demonstrating that action-related feedback is overweighted relative to incidental information would not justify concluding that delayed feedback is processed similarly to incidental information. The two biases could be comparable in magnitude, but entirely unrelated. Future work will be needed to resolve this

issue, perhaps using neuroimaging to identify whether Pavlovian circuits are recruited during delayed feedback processing.

Our study contributes to the understanding of temporal discounting by providing an alternative explanation for its prevalence and resistance to correction [11]. Unlike traditional intertemporal choice paradigms, which focus on known options and delays, our research highlights the role of learning in shaping impulsive behavior. If reinforcement loses efficacy with delay, immediate consequences would receive undue weight relative to delayed ones. This would lead to a discounting of delayed rewards, not because people do not like to wait, but because the delayed rewards are harder to learn. Thus, the learning bias that our study demonstrates may be a natural explanation for why impatience is so common, while the opposite, future bias, is so rare. Our results offer some evidence in this direction by showing a correlation of behavior in the learning task with typical measures of time discounting. Moreover, if time preferences are influenced by experience rather than inherent traits, discount rates should vary across domains within individuals. Empirical findings support this notion [71].

Our study was designed to rule out temporal discounting effects. First, our delays were on the order of seconds rather than hours or days, as typical in intertemporal choice tasks. Nonetheless, some studies have shown that temporal discounting can occur even with delays on the order of seconds [64,72]. More importantly, reward discounting can not explain the bias because all rewards were delivered at the end of the experiment. One might think that subjects could have willingly foregone monetary reward in order to receive good feedback a few seconds earlier. The high losses observed due to the bias make this implausible. In study 1 for example, subjects on average lost €1.73 in incongruent and €0.64 in congruent trials, meaning that the willingness to pay for good early feedback must have been at least €1.09 to explain the bias in

702 this manner ($t(86) = -6.75, p < 0.001$). Finally, the memory tasks indicate that subjects indeed
703 misestimated total rewards as reflected in their choices.

704 Cognitive load is another potential explanation for learning frictions. Our task required
705 subjects to process feedback from two different choices at the same time, one delayed and one
706 immediate. This could explain why subjects learn incompletely, but it doesn't explain why they
707 favor the immediate feedback – cognitive load should affect the immediate and delayed feedback
708 equally. Furthermore, once a subject has learned one feedback component (e.g., the immediate
709 one) they should shift their attention and learn the other, leading to a decrease in bias over time.
710 This is the opposite of what we observed. One possibility is that people might be holding in mind
711 the immediate feedback in order to add it to the delayed one. This might explain why the
712 immediate feedback seems to be double counted. Cognitive load has been shown to increase
713 dwell time [73] and this might be one reason why we don't observe an association between dwell
714 time and the immediacy bias. In any case, future work could study our task in a less constrained
715 setting where immediate and delayed feedback are given in isolation [27].

716 The prevalence of myopic behavior has significant implications across various domains.
717 In financial decision-making, it can lead to inadequate investments in retirement savings and
718 education, as well as overspending and debt accumulation. Myopic behavior is also associated
719 with health-related choices, where immediate gratification often takes precedence over long-term
720 well-being. Skill development and learning are affected by temporal discounting, as individuals
721 may prioritize immediate enjoyment over the sustained effort required for skill acquisition. This
722 can impact academic and professional performance, personal goals, and overall life satisfaction.
723 This bias could also be relevant in the realm of social media. The focus on immediate rewards,
724 coupled with algorithmic promotion of viral content, can influence users' sharing and liking

behaviors, potentially shaping online discourse and content consumption. Addressing and mitigating the effects of delayed discounting may involve strategies such as enhancing the visibility of future rewards [22] and emphasizing learning through observation.

In summary, we have shown that people show a bias to learn more from immediate feedback than even slightly delayed feedback. This has major implications for how we evaluate people's choices. As opposed to ascribing impatient behavior to an unwillingness to wait, we argue that impatience may in part be due to delayed rewards receiving too little weight in learning, perhaps due to inattention to delayed feedback. This may explain why impatience and temporal inconsistency are such pervasive problems.

Materials and Methods

Ethics Statement

Study 1 was ruled as exempt by the joint Ethics Committee of the Faculty of Economics and Business Administration of Goethe University Frankfurt (GU) and the Gutenberg School of Management & Economics of the Faculty of Law, Management and Economics of Johannes Gutenberg University Mainz (JGU). Written informed consent was obtained prior to data collection.

Studies 2 & 3 were approved by the Behavioral and Social Sciences Institutional Review Board at the Ohio State University, protocol 2013B0583. Written informed consent was obtained prior to data collection.

Study 4 was ruled as exempt by the UCLA Office of the Human Research Protection Program. Written informed consent was obtained prior to data collection.

Experimental Paradigm

In Study 1, subjects had 10 seconds to make their decisions and lost 5 points if they ran out of time. In Study 2, subjects had 3 seconds and lost 5 points if they ran out of time. In Study 2, the feedback screen was presented for 2 seconds, and a randomly generated time interval between two and six seconds was added after each choice and feedback screen.

Study 1 was programmed using oTree [74]. Study 2 was programmed using jsPsych [75] and utilized webcam-based eye tracking using the Webgazer library [49] integrated in jsPsych [41].

At the end of each study, subjects completed a memory survey. We asked subjects to rank the stimuli in terms of total reward, to indicate whether each stimulus was ascending, descending or flat, and to estimate the average total reward for each stimulus (SI Methods). We only analyzed the memory surveys for Study 2, since the memory survey for Study 1 only asked the questions for a subset of the stimuli presented (SI Methods). For the ranking question we used Kendall tau distance between the true rank and the recalled rank. This measure counts the number of pairwise disagreements between two rankings. For the ascending vs. descending question, we computed the sum of the errors. If the subject answered that the immediate feedback was equal to the delayed feedback, we counted it as half an error. For the total reward question, we calculated the average error.

At the end of Study 1, subjects also completed a temporal discounting task and a working-memory n-back task [76].

In the temporal discounting task, subjects completed a hypothetical survey to assess their intertemporal preferences. In three series of questions, subjects were asked how they would decide between 100 euros today or x in one month; 100 euros today or x in six months; 100

euros in one month or x in six months. By increasing or decreasing x from question to question, we obtained indifference points in the range 100–132 euros for each time horizon.

In the n -back task, participants were presented with a sequence of stimuli, in this case numbers, one at a time. The task required participants to indicate whether the current stimulus matched the one presented n items back in the sequence. One sequence was a 2-back task and consisted of 48 stimuli of which 14 were target stimuli and one sequence was a 3-back task and consisted of 48 stimuli of which 16 were target stimuli.

We assessed performance on the n -back task using accuracy and d prime. Accuracy includes both correctly identified targets and correctly identified non-targets. D prime is a measure of sensitivity or discriminability used in signal detection theory [77] and is calculated as the difference in z -scored hit rates and false alarm rates. These measures were corrected to account for cases in which the hit rate equals 1 or false alarm rate equals 0. In order to correct the measures, we added 0.5 to the total number of hits and false alarms and added 1 to the total number of stimuli [78,79].

The 3-back task was not used for some subjects. In total, 87 subjects completed the 2-back task and 51 subjects completed the 3-back task. 1 subject was excluded from the analyses because their d prime measure was negative, indicating worse-than-chance behavior.

For Study 2 and 3, we excluded some participants based on their eye-tracking data. A short validation phase was presented after trials 21, 42, and 84. After 63 trials participants completed a new calibration and validation. If participants failed all 3 short validations, they were excluded. Each validation check consisted of 3 dots at different positions on the screen that participants had to fixate on. A participant failed the validation if they failed to fixate on any of the 3 dots, within the error margin.

794

795 **Modeling**

796 *Models*

797 For the differential learning model, the values for the immediate and delayed feedback
798 components were as follows:

799

$$V_{Ik+1}(s_k) = V_{Ik}(s_k) + \alpha^I(r_{Ik} - V_{Ik}(s_k)) \quad (1)$$

$$V_{Dk+1}(s_k) = V_{Dk}(s_k) + \alpha^D(r_{Dk} - V_{Dk}(s_k)) \quad (2)$$

The total value of a stimulus was the sum of the immediate and delayed values.

$$V_k(s_k) = V_{Ik}(s_k) + V_{Dk}(s_k) \quad (3)$$

800 Each value was initialized to 0 at $k = 0$.

801 For the baseline learning model, the learning rates for the immediate and delayed
802 components were the same ($\alpha^I = \alpha^D = \alpha$).

803 For the differential weight model, the learning rates for the immediate and delayed
804 components were the same but the estimated value of the delayed feedback was weighted less
805 than the estimated value of the immediate feedback. The total value of a stimulus was the sum of
806 the immediate and delayed values:

$$V_k(s_k) = V_{Ik}(s_k) + \gamma V_{Dk}(s_k) \quad (4)$$

807 where γ is the discount factor on the estimated value of the delayed component.

808 For the differential learning and weight model, the learning rates and choice weights for
809 the immediate and delayed feedback were allowed to be different. This model was a combination
810 of the differential learning model and the differential weight model.

The decision process itself was described by the DDM [16,80] with one boundary for correct responses and one boundary for incorrect responses. In each trial, the drift rate was defined as the difference between the values from the RL model. When the difference was high, the drift rate was higher, leading to more accurate and faster responses. On each trial, the drift rate was defined as:

$$v \sim d(V_{k,correct}(S_k) - V_{k,incorrect}(S_k)) \quad (5)$$

To compare models, we used the Widely Applicable Information Criterion (WAIC) [81]. This criterion is especially useful when the posterior distribution is not Normal. We computed WAIC for each subject and each model.

Model comparison provides only relative performance among models. To check whether the RL model captured the data well, we conducted a posterior predictive check. For each subject, we used the mean of the posterior distribution to generate 100 simulated choices and RTs using their actual sequence of trials. We computed choice accuracy at three levels: (1) at the trial-level when averaging across all subjects (Fig. 5A, B), (2) at the subject-level when averaging across all trials (Fig. S12), (3) overall level when averaging across both trials and subjects (Fig. S13) and response times across both trials and subjects (Fig. S14) [80]. We excluded trials where the two options were identical.

Priors

We chose weakly informative priors so as to have minimal influence on the posterior [82].

$$\alpha^l \sim B(1,1), 0 \leq \alpha^l \leq 1 \quad (6)$$

$$\alpha^D \sim B(1,1), 0 \leq \alpha^D \leq 1 \quad (7)$$

$$a \sim N(2,1), a \geq 0 \quad (8)$$

$$t \sim U(0, \min RT) \quad (9)$$

$$d \sim N(0, 2), d \geq 0 \quad (10)$$

$$\gamma \sim B(1,1), 0 \leq \gamma \leq 1 \quad (11)$$

832

833

834 *Fitting procedure*

835 We used Stan to find the best fitting parameters for each subject separately, using the Monte-
 836 Carlo Markov Chain sampling method. For each subject, four chains were run in parallel. Each
 837 chain consisted of 4,000 samples out of which 2,000 were warm-up samples. We computed R-
 838 hat of all parameters to assess model convergence. This method compares the between- and
 839 within-chain estimates for model parameters. The maximum R-hat was 1.01 indicating model
 840 convergence [82].

841

842 *Parameter recovery*

843 To test whether parameters of the models could be recovered well, we simulated one dataset for
 844 176 subjects using the mean posterior of the fitted parameter values and the experimental trials
 845 encountered by each subject in the task. We then fit the model on this simulated data. We found
 846 large and significant correlations between the true and fitted parameter values and no significant
 847 bias in the recovery of the model parameters (Fig. S12).

848

849 *Model recovery*

To test whether the differential learning model and the differential weight model could be distinguished from one another, we examined model recovery. We simulated one dataset for 176 participants using the mean posteriors of the fitted parameter values and the experimental trials encountered by each participant in the task for each model. We then fit each model on these simulated datasets and compared the model fits using WAIC to determine whether the model used to generate the data was also the best-fitting model. To check whether more trials leads to better model recovery, we also simulated datasets for each model using double the number of trials.

The model recovery analyses show that the differential weight model tends to mimic the differential learning model even when the differential learning model generated the data, but not the other way around (Simulated: Differential Learning Model, Fitted: 53% Differential Learning Model, 47% Differential Weight Model; Simulated: Differential Weight Model, Fitted: 72% Weight Model, 28% Differential Learning Model; Fig. S22). This is true even when using double the number of trials in the experiment (Simulated: Differential Learning Model, Fitted: 65% Differential Learning Model, 35% Differential Weight Model; Simulated: Differential Weight Model, Fitted: 71% Weight Model, 29% Differential Learning Model; Fig. S22).

Acknowledgments

We thank Adam Foa, Malia Young, Sumedha Goyal, and Walker Daniel for research assistance. National Science Foundation grant 2333979 (IK).

Data and materials availability:

The code to replicate the analyses, tables and figures is available on OSF at the following link:

873 https://osf.io/rnwxb/overview?view_only=f4220b32affa45eb9e30d514030f6811

874 Analyses were conducted using R and stan.

875

876 **Preregistrations:**

877 The preregistration for Study 2 is available at OSF:

878 https://osf.io/mkggy/overview?view_only=de762d7b32184b16b6761c07ee856da8

879 for Study 3 is available at OSF:

880 https://osf.io/37qa2/overview?view_only=15f9c403b0ef4c2abf8eefca68d45421

881 and for Study 4 is available at OSF:

882 https://osf.io/tuv48/overview?view_only=68540a1dab854d9ba9d329860041bede

883

884 **Author Contributions:**

885 Conceptualization: DP

886 Methodology: MC, DP, IK

887 Investigation: MC, DP

888 Visualization: MC, DP

889 Funding acquisition: IK, DP

890 Project administration: IK, DP

891 Supervision: IK

892 Writing – original draft: MC, DP, IK

893 Writing – review & editing: MC, DP, IK

894

895

896

897

898

899

900

References

1. Doll BB, Simon DA, Daw ND. The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*. 2012 Dec 1;22(6):1075–81. doi:10.1016/j.conb.2012.08.003
2. Glimcher PW. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *PNAS*. 2011 Sep 13;108(Supplement 3):15647–54. doi:10.1073/pnas.1014269108 PubMed PMID: 21389268.
3. Daw ND, Doya K. The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*. 2006 Apr;16(2):2. doi:10.1016/j.conb.2006.03.006
4. Sutton RS, Barto AG. *Reinforcement learning: An introduction*, 2nd ed. Cambridge, MA, US: The MIT Press; 2018. xxii, 526 p. (Reinforcement learning: An introduction, 2nd ed).
5. Lee AS, Duman RS, Pittenger C. A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. *PNAS*. 2008 Nov 4;105(44):17163–8. doi:10.1073/pnas.0807749105
6. O’Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron*. 2003 Apr 24;38(2):329–37. doi:10.1016/S0896-6273(03)00169-7
7. Schultz W. Subjective neuronal coding of reward: temporal value discounting and risk. *European Journal of Neuroscience*. 2010;31(12):2124–35. doi:10.1111/j.1460-9568.2010.07282.x
8. Gershman SJ. The Successor Representation: Its Computational Logic and Neural Substrates. *J Neurosci*. 2018 Aug 15;38(33):7193–200. doi:10.1523/JNEUROSCI.0151-18.2018 PubMed PMID: 30006364.

- 938 9. Rubin JE, Vich C, Clapp M, Noneman K, Verstynen T. The credit assignment problem in
939 cortico-basal ganglia-thalamic networks: A review, a problem and a possible solution. *Eur J*
940 *Neurosci.* 2021 Apr;53(7):2234–53. doi:10.1111/ejn.14745 PubMed PMID: 32302439.
- 941 10. Witkowski PP, Park SA, Boorman ED. Neural mechanisms of credit assignment for inferred
942 relationships in a structured world. *Neuron.* 2022 Aug 17;110(16):2680-2690.e9.
943 doi:10.1016/j.neuron.2022.05.021
- 944 11. Zauberman G, Kim BK, Malkoc SA, Bettman JR. Discounting Time and Time Discounting:
945 Subjective Time Perception and Intertemporal Preferences. *Journal of Marketing Research.*
946 2009 Aug 1;46(4):543–56. doi:10.1509/jmkr.46.4.543
- 947 12. Frederick S, Loewenstein G. Time Discounting and Time Preference: A Critical Review.
948 *Journal of Economic Literature.* 2002.
- 949 13. Ericson KM, Laibson D. Intertemporal choice. In: *Handbook of Behavioral Economics:*
950 *Applications and Foundations 1* [Internet]. North-Holland; 2019 [cited 2025 Nov 25]. p. 1–
951 67. Available from:
952 <https://www.sciencedirect.com/science/chapter/handbook/abs/pii/S2352239918300253>
953 doi:10.1016/bs.hesbe.2018.12.001
- 954 14. Kable JW. Valuation, Intertemporal Choice, and Self-Control. In: Glimcher PW, Fehr E,
955 editors. *Neuroeconomics.* Academic Press; 2014. p. 173–92. doi:10.1016/B978-0-12-
956 416008-8.00010-3
- 957 15. Dai J, Busemeyer JR. A probabilistic, dynamic, and attribute-wise model of intertemporal
958 choice. *Journal of Experimental Psychology: General.* 2014;143(4):1489–514.
959 doi:10.1037/a0035976
- 960 16. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of
961 value in simple choice. *Nat Neurosci.* 2010 Oct;13(10):1292–8. doi:10.1038/nn.2635
- 962 17. Shevlin BRK, Krajbich I. Attention as a source of variability in decision-making: Accounting
963 for overall-value effects with diffusion models. *Journal of Mathematical Psychology.*
964 2021;105. doi:10.1016/j.jmp.2021.102594
- 965 18. Cockburn J, Collins AGE, Frank MJ. A Reinforcement Learning Mechanism Responsible for
966 the Valuation of Free Choice. *Neuron.* 2014 Aug 6;83(3):551–7.
967 doi:10.1016/j.neuron.2014.06.035 PubMed PMID: 25066083.
- 968 19. Murty VP, DuBrow S, Davachi L. The simple act of choosing influences declarative memory.
969 *J Neurosci.* 2015 Apr 22;35(16):6255–64. doi:10.1523/JNEUROSCI.4181-14.2015 PubMed
970 PMID: 25904779; PubMed Central PMCID: PMC4405547.

- 971 20. Chambon V, Théro H, Vidal M, Vandendriessche H, Haggard P, Palminteri S. Information
972 about action outcomes differentially affects learning from self-determined versus imposed
973 choices. *Nat Hum Behav.* 2020 Oct;4(10):1067–79. doi:10.1038/s41562-020-0919-5
- 974 21. Pupillo F, Bruckner R. Signed and unsigned effects of prediction error on memory: Is it a
975 matter of choice? *Neuroscience & Biobehavioral Reviews.* 2023 Oct 1;153:105371.
976 doi:10.1016/j.neubiorev.2023.105371
- 977 22. Chen F, Zheng J, Wang L, Krajbich I. Attribute latencies causally shape intertemporal
978 decisions. *Nat Commun.* 2024 Apr 5;15(1):2948. doi:10.1038/s41467-024-46657-2
- 979 23. Maier SU, Raja Beharelle A, Polanía R, Ruff CC, Hare TA. Dissociable mechanisms govern
980 when and how strongly reward attributes affect decisions. *Nat Hum Behav.* 2020
981 Sep;4(9):949–63. doi:10.1038/s41562-020-0893-y
- 982 24. Sullivan NJ, Huettel SA. Healthful choices depend on the latency and rate of information
983 accumulation. *Nat Hum Behav.* 2021 Dec;5(12):1698–706. doi:10.1038/s41562-021-01154-
984 0
- 985 25. Weber EU, Johnson EJ, Milch KF, Chang H, Brodscholl JC, Goldstein DG. Asymmetric
986 Discounting in Intertemporal Choice: A Query-Theory Account. *Psychol Sci.* 2007 Jun
987 1;18(6):516–23. doi:10.1111/j.1467-9280.2007.01932.x
- 988 26. Zhang Z, Wang S, Good M, Hristova S, Kayser AS, Hsu M. Retrieval-constrained valuation:
989 Toward prediction of open-ended decisions. *Proceedings of the National Academy of*
990 *Sciences.* 2021 May 18;118(20):e2022685118. doi:10.1073/pnas.2022685118
- 991 27. Foerde K, Shohamy D. Feedback timing modulates brain systems for learning in humans. *J*
992 *Neurosci.* 2011 Sep 14;31(37):13157–67. doi:10.1523/JNEUROSCI.2701-11.2011 PubMed
993 PMID: 21917799; PubMed Central PMCID: PMC3328791.
- 994 28. Knowlton BJ, Mangels JA, Squire LR. A Neostriatal Habit Learning System in Humans.
995 *Science.* 1996 Sep 6;273(5280):1399–402. doi:10.1126/science.273.5280.1399 PubMed
996 PMID: 8703077.
- 997 29. Foerde K, Race E, Verfaellie M, Shohamy D. A Role for the Medial Temporal Lobe in
998 Feedback-Driven Learning: Evidence from Amnesia. *J Neurosci.* 2013 Mar 27;33(13):5698–
999 704. doi:10.1523/JNEUROSCI.5217-12.2013 PubMed PMID: 23536083.
- 1000 30. Pavlov IP. Conditioned reflexes: an investigation of the physiological activity of the cerebral
1001 cortex. Oxford, England: Oxford Univ. Press; 1927. xv, 430 p. (Conditioned reflexes: an
1002 investigation of the physiological activity of the cerebral cortex).
- 1003 31. Bangasser DA, Waxler DE, Santollo J, Shors TJ. Trace conditioning and the hippocampus:
1004 the importance of contiguity. *J Neurosci.* 2006 Aug 23;26(34):8702–6.

doi:10.1523/JNEUROSCI.1742-06.2006 PubMed PMID: 16928858; PubMed Central PMCID: PMC3289537.

32. Rescorla R, Wagner A. A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement. undefined [Internet]. 1972 [cited 2021 Apr 23]. Available from: /paper/A-theory-of-Pavlovian-conditioning-%3A-Variations-in-Rescorla/afaf65883ff75cc19926f61f181a687927789ad1

33. Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput.* 2012 May;24(5):1186–229. doi:10.1162/NECO_a_00270 PubMed PMID: 22295983.

34. Konovalov A, Krajbich I. Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nat Commun.* 2016 Aug 11;7(1):12438. doi:10.1038/ncomms12438

35. Fontanesi L, Gluth S, Spektor MS, Rieskamp J. A reinforcement learning diffusion decision model for value-based decisions. *Psychon Bull Rev.* 2019 Aug 1;26(4):1099–121. doi:10.3758/s13423-018-1554-2

36. Pedersen ML, Frank MJ. Simultaneous Hierarchical Bayesian Parameter Estimation for Reinforcement Learning and Drift Diffusion Models: a Tutorial and Links to Neural Data. *Comput Brain Behav.* 2020 Dec;3(4):458–71. doi:10.1007/s42113-020-00084-w PubMed PMID: 35128308; PubMed Central PMCID: PMC8811713.

37. Shevlin BRK, Smith SM, Hausfeld J, Krajbich I. High-value decisions are fast and accurate, inconsistent with diminishing value sensitivity. *Proc Natl Acad Sci U S A.* 2022 Feb 8;119(6):e2101508119. doi:10.1073/pnas.2101508119 PubMed PMID: 35105801; PubMed Central PMCID: PMC8832986.

38. Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, Behrens TEJ. Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci.* 2012 Mar;15(3):470–6. doi:10.1038/nn.3017

39. Pirrone A, Azab H, Hayden BY, Stafford T, Marshall JAR. Evidence for the speed–value trade-off: Human and monkey decision making is magnitude sensitive. *Decision.* 2018;5(2):129–42. doi:10.1037/dec0000075

40. Shenhav A, Karmarkar UR. Dissociable components of the reward circuit are involved in appraisal versus choice. *Sci Rep.* 2019 Feb 13;9(1):1958. doi:10.1038/s41598-019-38927-7

41. Yang X, Krajbich I. Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making.* 2021 Nov;16(6):1485–505. doi:10.1017/S1930297500008512

- 1038 42. Busemeyer JR, Townsend JT. Decision field theory: a dynamic-cognitive approach to
1039 decision making in an uncertain environment. *Psychol Rev.* 1993 Jul;100(3):432–59.
1040 doi:10.1037/0033-295x.100.3.432 PubMed PMID: 8356185.
- 1041 43. Diederich A. Dynamic stochastic models for decision making under time constraints.
1042 *Journal of Mathematical Psychology.* 1997;41(3):260–74. doi:10.1006/jmps.1997.1167
- 1043 44. Roe RM, Busemeyer JR, Townsend JT. Multialternative decision field theory: a dynamic
1044 connectionist model of decision making. *Psychol Rev.* 2001 Apr;108(2):370–92.
1045 doi:10.1037/0033-295x.108.2.370 PubMed PMID: 11381834.
- 1046 45. Konovalov A, Krajbich I. Revealed strength of preference: Inference from response times.
1047 Judgment and decision making [Internet]. 2019 Jul [cited 2025 Nov 25];14(4). Available
1048 from: [https://par.nsf.gov/biblio/10288586-revealed-strength-preference-inference-from-](https://par.nsf.gov/biblio/10288586-revealed-strength-preference-inference-from-response-times)
1049 [response-times](https://par.nsf.gov/biblio/10288586-revealed-strength-preference-inference-from-response-times)
- 1050 46. Ratcliff R, McKoon G. The Diffusion Decision Model: Theory and Data for Two-Choice
1051 Decision Tasks. *Neural Comput.* 2008 Apr;20(4):873–922. doi:10.1162/neco.2008.12-06-
1052 420 PubMed PMID: 18085991; PubMed Central PMCID: PMC2474742.
- 1053 47. Polanía R, Krajbich I, Grueschow M, Ruff CC. Neural oscillations and synchronization
1054 differentially support evidence accumulation in perceptual and value-based decision
1055 making. *Neuron.* 2014 May 7;82(3):709–20. doi:10.1016/j.neuron.2014.03.014 PubMed
1056 PMID: 24811387.
- 1057 48. Cavanagh JF, Wiecki TV, Kochar A, Frank MJ. Eye Tracking and Pupillometry are Indicators of
1058 Dissociable Latent Decision Processes. *J Exp Psychol Gen.* 2014 Aug;143(4):1476–88.
1059 doi:10.1037/a0035813 PubMed PMID: 24548281; PubMed Central PMCID: PMC4114997.
- 1060 49. Papoutsaki A. Scalable Webcam Eye Tracking by Learning from User Interactions. In:
1061 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in
1062 Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery;
1063 2015 [cited 2025 Nov 25]. p. 219–22. (CHI EA '15). Available from:
1064 <https://doi.org/10.1145/2702613.2702627> doi:10.1145/2702613.2702627
- 1065 50. Semmelmann K, Weigelt S. Online webcam-based eye tracking in cognitive science: A first
1066 look. *Behav Res Methods.* 2018 Apr;50(2):451–65. doi:10.3758/s13428-017-0913-7
1067 PubMed PMID: 28593605.
- 1068 51. Herrnstein RJ, Prelec D. Melioration: A Theory of Distributed Choice. *The Journal of*
1069 *Economic Perspectives.* 1991;5(3):137–56.
- 1070 52. Herrnstein RJ. Behavior, Reinforcement and Utility. *Psychol Sci.* 1990 Jul 1;1(4):217–24.
1071 doi:10.1111/j.1467-9280.1990.tb00203.x

- 1072 53. Herrnstein RJ. Experiments on stable sub-optimality in individual behavior. American
1073 Economic Review. 1991;360–4.
- 1074 54. Heyman GM, Herrnstein RJ. More on Concurrent Interval-Ratio Schedules: A Replication
1075 and Review. Journal of the Experimental Analysis of Behavior. 1986;46(3):331–51.
1076 doi:10.1901/jeab.1986.46-331
- 1077 55. Vaughan Jr. W. Melioration, Matching, and Maximization. Journal of the Experimental
1078 Analysis of Behavior. 1981;36(2):141–9. doi:10.1901/jeab.1981.36-141
- 1079 56. Herrnstein RJ, Loewenstein GF, Prelec D, Vaughan Jr. W. Utility maximization and
1080 melioration: Internalities in individual choice. Journal of Behavioral Decision Making.
1081 1993;6(3):149–85. doi:10.1002/bdm.3960060302
- 1082 57. Herrnstein, R.J. The Matching Law: Papers in Psychology and Economics. Rachlin H, D. I.
1083 Laibson, editors. Harvard University Press; 1997.
- 1084 58. Prelec D. Consuming at the Wrong Rate: Lessons from the Harvard Game. In: Sustainable
1085 Consumption: Multi-disciplinary Perspectives In Honour of Professor Sir Partha Dasgupta |
1086 Oxford Academic [Internet]. Oxford University Press; 2014 [cited 2025 Nov 25]. Available
1087 from: [https://academic.oup.com/book/32654/chapter-](https://academic.oup.com/book/32654/chapter-abstract/270592248?redirectedFrom=fulltext)
1088 [abstract/270592248?redirectedFrom=fulltext](https://academic.oup.com/book/32654/chapter-abstract/270592248?redirectedFrom=fulltext)
- 1089 59. Gureckis TM, Love BC. Short-term gains, long-term pains: How cues about state aid
1090 learning in dynamic environments. Cognition. 2009 Dec;113(3):293–313.
1091 doi:10.1016/j.cognition.2009.03.013
- 1092 60. Neth H, Sims CR, Gray WD. Melioration Dominates Maximization: Stable Suboptimal
1093 Performance Despite Global Feedback. 6.
- 1094 61. Tunney RJ, Shanks DR. A re-examination of melioration and rational choice. J Behav Decis
1095 Making. 2002 Oct;15(4):291–311. doi:10.1002/bdm.415
- 1096 62. Balasubramani PP, Diaz-Delgado J, Grennan G, Alim F, Zafar-Khan M, Maric V, et al. Distinct
1097 neural activations correlate with maximization of reward magnitude versus frequency.
1098 Cereb Cortex. 2023 May 9;33(10):6038–50. doi:10.1093/cercor/bhac482 PubMed PMID:
1099 36573422; PubMed Central PMCID: PMC10422923.
- 1100 63. Sims CR, Neth H, Jacobs RA, Gray WD. Melioration as rational choice: Sequential decision
1101 making in uncertain environments. Psychological Review. 2012;120(1):139.
1102 doi:10.1037/a0030850
- 1103 64. Fiorillo CD, Newsome WT, Schultz W. The temporal precision of reward prediction in
1104 dopamine neurons. Nature Neuroscience. 2008 Aug;11(8):8. doi:10.1038/nn.2159

- 1105 65. Kobayashi S, Schultz W. Influence of Reward Delays on Responses of Dopamine Neurons. *J*
 1106 *Neurosci.* 2008 Jul 30;28(31):7837–46. doi:10.1523/JNEUROSCI.1600-08.2008 PubMed
 1107 PMID: 18667616.
- 1108 66. Foerde K, Shohamy D. The role of the basal ganglia in learning and memory: insight from
 1109 Parkinson's disease. *Neurobiol Learn Mem.* 2011 Nov;96(4):624–36.
 1110 doi:10.1016/j.nlm.2011.08.006 PubMed PMID: 21945835; PubMed Central PMCID:
 1111 PMC3772079.
- 1112 67. Lighthall NR, Pearson JM, Huettel SA, Cabeza R. Feedback-Based Learning in Aging:
 1113 Contributions and Trajectories of Change in Striatal and Hippocampal Systems. *J Neurosci.*
 1114 2018 Sep 26;38(39):8453–62. doi:10.1523/JNEUROSCI.0769-18.2018 PubMed PMID:
 1115 30120208.
- 1116 68. Bakkour A, Palombo DJ, Zylberberg A, Kang YH, Reid A, Verfaellie M, et al. The
 1117 hippocampus supports deliberation during value-based decisions. Kahnt T, Frank MJ,
 1118 Fellows LK, Gluth S, editors. *eLife.* 2019 Jul 3;8:e46080. doi:10.7554/eLife.46080
- 1119 69. Newell BR, Dunn JC, Kalish M. Systems of category learning: Fact or fantasy? In: *The*
 1120 *psychology of learning and motivation: Advances in research and theory*, Vol. 54. San
 1121 Diego, CA, US: Elsevier Academic Press; 2011. p. 167–215. (The psychology of learning and
 1122 motivation).
- 1123 70. Dorfman HM, Gershman SJ. Controllability governs the balance between Pavlovian and
 1124 instrumental action selection. *Nat Commun.* 2019 Dec 20;10(1):5826. doi:10.1038/s41467-
 1125 019-13737-7
- 1126 71. Gabaix X, Laibson D. Myopia and Discounting [Working Paper] [Internet]. National Bureau
 1127 of Economic Research; 2017 [cited 2025 Nov 25]. (Working Paper Series). Available from:
 1128 <https://www.nber.org/papers/w23254> doi:10.3386/w23254
- 1129 72. Gregorios-Pippas L, Tobler PN, Schultz W. Short-Term Temporal Discounting of Reward
 1130 Value in Human Ventral Striatum. *Journal of Neurophysiology.* 2009 Mar 1;101(3):1507–23.
 1131 doi:10.1152/jn.90730.2008
- 1132 73. Orquin JL, Mueller Loose S. Attention and choice: A review on eye movements in decision
 1133 making. *Acta Psychologica.* 2013 Sep 1;144(1):190–206. doi:10.1016/j.actpsy.2013.06.003
- 1134 74. Chen DL, Schonger M, Wickens C. oTree—An open-source platform for laboratory, online,
 1135 and field experiments. *Journal of Behavioral and Experimental Finance.* 2016 Mar 1;9:88–
 1136 97. doi:10.1016/j.jbef.2015.12.001
- 1137 75. de Leeuw JR. jsPsych: a JavaScript library for creating behavioral experiments in a Web
 1138 browser. *Behav Res Methods.* 2015 Mar;47(1):1–12. doi:10.3758/s13428-014-0458-y
 1139 PubMed PMID: 24683129.

- 1140 76. Kirchner WK. Age differences in short-term retention of rapidly changing information. J Exp
1141 Psychol. 1958 Apr;55(4):352–8. doi:10.1037/h0043688 PubMed PMID: 13539317.
- 1142 77. Wixted JT. The forgotten history of signal detection theory. Journal of Experimental
1143 Psychology: Learning, Memory, and Cognition. 2020;46(2):201–33.
1144 doi:10.1037/xlm0000732
- 1145 78. Hautus MJ. Corrections for extreme proportions and their biasing effects on estimated
1146 values of d' . Behavior Research Methods, Instruments & Computers. 1995;27(1):46–51.
1147 doi:10.3758/BF03203619
- 1148 79. Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behavior
1149 Research Methods, Instruments & Computers. 1999;31(1):137–49.
1150 doi:10.3758/BF03207704
- 1151 80. Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion Decision Model: Current Issues and
1152 History. Trends Cogn Sci. 2016 Apr;20(4):260–81. doi:10.1016/j.tics.2016.01.007 PubMed
1153 PMID: 26952739; PubMed Central PMCID: PMC4928591.
- 1154 81. Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable
1155 Information Criterion in Singular Learning Theory. J Mach Learn Res. 2010
1156 Dezember;11:3571–94.
- 1157 82. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis,
1158 Third Edition. CRC Press; 2013. 677 p.

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178	Supporting Information
1179	
1180	Supplementary Text
1181	
1182	S1 Text. Supplementary methods and results. Contains methods and additional results for Studies
1183	1–4, including behavioral analyses, response times, gaze data, model fits, and memory and
1184	preference measures.
1185	
1186	Supplementary Figures
1187	
1188	S1 Fig. Behavioral bias at the subject level for Study 1 Colors.
1189	S2 Fig. Behavioral bias at the subject level for Study 2 Patterns.
1190	S3 Fig. Behavioral bias at the subject level for Study 3 Patterns Reversed Feedback Position.
1191	S4 Fig. Behavioral bias at the subject level for Study 4 In-lab Eye-tracking.
1192	S5 Fig. RT indifference point.
1193	S6 Fig. Attention to feedback screen for each subject for Studies 2 and 3 Patterns.
1194	S7 Fig. Attention to feedback screen for each subject for Study 4 In-lab Eye-tracking Study.
1195	S8 Fig. Attention to the choice screen for each subject for Studies 2 and 3 Patterns.
1196	S9 Fig. Attention to the choice screen for each subject for Study 4 In-lab Eye-tracking Study.
1197	S10 Fig. Behavioral bias and attention.
1198	S11 Fig. Predicted aggregate behavioral bias.
1199	S12 Fig. Predicted and observed accuracy at the subject level.
1200	S13 Fig. Observed and predicted experiment accuracy.
1201	S14 Fig. Observed and predicted value difference (VD) and overall value (OV) effects on RT.
1202	S15 Fig. RL parameters recovery.
1203	S16 Fig. Behavioral bias for reversed feedback position.
1204	S17 Fig. Behavioral bias for in-lab eye-tracking study.
1205	S18 Fig. Value difference (VD) and overall value (OV) effects on response times (RT) for
1206	reversed feedback position condition.
1207	S19 Fig. Value difference (VD) and overall value (OV) effects on response times (RT) for in-lab
1208	eye-tracking study.
1209	S20 Fig. Attention and behavioral bias for reversed feedback position.
1210	S21 Fig. Attention and behavioral bias for in-lab eye-tracking data.
1211	S22 Fig. Model recovery.
1212	S23 Fig. Model simulations with double the number of trials compared to the experiment.
1213	S24 Fig. Behavioral bias over time in the passive learning condition.
1214	
1215	Supplementary Tables
1216	
1217	S1 Table. Type of option effect on choice.
1218	S2 Table. Type of reward effect on choice.
1219	S3 Table. Value difference (VD) and overall value effects (OV) on response time.
1220	S4 Table. Value difference (VD) and overall value (OV) effects by type of reward on response
1221	time.
1222	S5 Table. Value difference effects between descending and ascending options on response time.
1223	S6 Table. Attention and reward feedback.

1224	S7 Table. Attention and model predictions.
1225	S8 Table. Dwell proportion difference on choice.
1226	S9 Table. Choice and memory errors.
1227	S10 Table. Bias and memory errors.
1228	S11 Table. Bias and n-back working memory task performance.
1229	S12 Table. Bias and intertemporal preferences.
1230	S13 Table. Bias and intertemporal preferences in-lab eye-tracking study.
1231	S14 Table. Type of option effect on choice by feedback position.
1232	S15 Table. Type of choice set effect on error rate by feedback position.
1233	S16 Table. Type of reward effect on choice by feedback position.
1234	S17 Table. Value difference (VD) and overall value effects (OV) on response time by feedback
1235	position.
1236	S18 Table. Value difference (VD) and overall value (OV) effects by type of reward on response
1237	time by feedback position.
1238	S19 Table. Value difference effects between descending and ascending options on response time
1239	by feedback position.
1240	
1241	Supplementary Materials
1242	
1243	S2 Text. Supplementary materials. Contains task instructions and comprehension quizzes for the
1244	Colors Task and Patterns Tasks.
1245	