The Rational Irrational: Better Learners Show Stronger Reward Frequency Biases

Mianzhi Hu[1] & Darrell A. Worthy[1]

[1] *Department of Psychological & Brain Sciences, Texas A&M University, College Station, TX*

Corresponding Author:
Mianzhi Hu
Ph.D. Student
Texas A&M University
4235 TAMU
College Station, TX 77845-4235
rudolfhu@tamu.edu

Abstract

Frequency effects, defined as a bias toward more frequently rewarded but less valuable options, have traditionally been viewed as maladaptive decision-making deficits. In the present study, we used a within-subject design in which participants completed a four-option reinforcement learning task twice, once under a baseline condition and once with a reward frequency manipulation, to test whether better baseline learning predicts greater or lesser susceptibility to frequency-based biases. Participants were first trained on two fixed option pairs and then transferred their knowledge to novel pairings in a testing phase. Across conditions, higher training accuracy generally predicted higher testing accuracy, with one critical exception: on trials where a more valuable option was pitted against a more frequently rewarded but less valuable alternative, participants with higher training accuracy exhibited a stronger bias toward the more frequent option. Moreover, baseline optimal choice rates in these specific trials were unrelated to—and even slightly negatively correlated with—optimal choice rates under the frequency condition. Computational modeling further showed that participants with better baseline learning performance were better fit by frequency-sensitive models in the frequency condition and they weighed frequency-based processing more heavily than value-based processing. Overall, these findings suggest that frequency effects, rather than reflect flawed learning, manifest more strongly in individuals with better baseline learning performance. This seemingly irrational bias may, under conditions of uncertainty, represent a flexible, adaptive strategy that emerges among the best learners when value-based approaches are costly or unreliable.

Introduction

Would you choose to dine at a decent local restaurant you know well, or take a chance on a new one with a higher Yelp rating? Would you stick with your usual route to work, or try a newly discovered shortcut that might be faster? Would you keep using a familiar app, or switch to an unfamiliar one that promises better features? If you tend to choose the former option, you are demonstrating what cognitive psychologists call a reward-frequency-based bias, or frequency effect (Don et al., 2019; Don & Worthy, 2022). Frequency effect refers to people's tendency to favor options that have yielded more frequent rewards, even if they offer lower average long-term payoffs (Don et al., 2019; Don & Worthy, 2022; Hu et al., 2025). This tendency can push people towards suboptimal decisions, such as dining at worse restaurants, taking longer commutes, or settling for less efficient tools. Yet under real-world constraints, where value-based calculations can be overly costly or practically impossible, sticking with familiar, reliably rewarding options may, in fact, serve as an adaptive heuristic (Gigerenzer & Gaissmaier, 2011). The current study seeks to examine whether such frequency effects reflect flawed value learning or instead represent a fundamental component of adaptive decision-making, thereby emerging more strongly in individuals with stronger learning and value integration skills.

Frequency monitoring underlies many classic behavioral decision-making heuristics (Gigerenzer & Gaissmaier, 2011), such as tallying (McCammon & Hägeli, 2007), fluency (Schooler & Hertwig, 2005), and mere-exposure effects (Fang et al., 2007; Zajonc et al., 1974). Frequency has been posited as one of the most basic ways the human brain encodes statistical information (Cosmides & Tooby, 1996; Gigerenzer, 1996; Obrecht et al., 2009). In the context of reinforcement learning (RL), however, a bias toward more frequently rewarded options despite their lower overall value has been traditionally interpreted as a decision-making deficit, particularly associated with severe psychiatric conditions (Ritter et al., 2004; Shurman et al., 2005) or brain injuries (Bechara et al., 1994, 1996). Supporting this view, frequency effects have been found to be more pronounced in populations with reduced cognitive functioning, including older adults (Don et al., 2022), individuals with substance use disorders (Kim et al., 2011; Stout et al., 2004), and those with developmental disorders (Sallum et al., 2013; Toplak et al., 2005).

Yet more recent findings challenge this deficit-based interpretation by showing that healthy participants without neuropsychological impairments would nevertheless show frequency effects and favor the more frequently rewarded option over objectively more valuable alternatives (Horstmann et al., 2012; Kumar et al., 2019; Lin et al., 2007, 2013; Steingroever et al., 2013). Moreover, these effects do not reliably distinguish between healthy and clinical populations (Kumar et al., 2019; North & O'Carroll, 2001; Upton et al., 2012; Wilder et al., 1998). Replications have been reported across several paradigms, including the Iowa Gambling Task (IGT) (Kumar et al., 2019; Lin et al., 2013; Steingroever et al., 2013; Upton et al., 2012), a modified version of IGT (Chiu & Lin, 2007; Lin et al., 2007), the Soochow Gambling Task (SGT) (Chiu et al., 2008; Lin et al., 2009; Upton et al., 2012), and other four-option RL tasks (Don et al., 2019; Don & Worthy, 2022; Hu et al., 2025). While there may be additional task-specific factors contributing to preferences for more frequently rewarded options in some of these paradigms (e.g., reward magnitude in the IGT or SGT), converging evidence generally suggests that reward frequency may represent a fundamental component of value learning rather than a pathological byproduct.

Why, then, do people show frequency effects? From the perspective of bounded rationality, humans make decisions under constraints of time, information, and cognitive resources, which often obscure the true structure of the environment (Gigerenzer & Brighton, 2009; Gigerenzer & Gaissmaier, 2011). Under such conditions, complex value-based computations may need to give way to simpler heuristics to achieve an effort-accuracy balance. In some cases, these heuristics can even lead to better decision outcomes compared to strictly value-based strategies (Gigerenzer & Brighton, 2009; Gigerenzer & Goldstein, 1996). Consistent with this view, research has shown that these mental shortcuts are especially favored when the decision-making environment is complex (Hogarth & Karelaia, 2007; Kool et al., 2017; Payne et al., 1988), time is limited (Rieskamp & Hoffrage, 2008; Wu et al., 2022), or information is incomplete (Hertwig & Erev, 2009; Yechiam & Busemeyer, 2005). Frequency effects may reflect one such adaptive shortcut. Previous studies have found that frequency effects emerge primarily when value differences between options are small, whereas when value differences are large and clear, reinforcement frequency exerts little influence on people's choices (Don et al., 2019; Don & Worthy, 2022). A recent study (Hu et al., 2025) further demonstrates that preference for the more frequently rewarded option increases proportionally with the level of environmental uncertainty. When outcome variance (i.e., uncertainty) was low, participants preferred the more valuable option; when variance was moderate, no clear preference emerged; and when variance was high, the frequently rewarded option was favored. Together, these findings suggest that frequency effects may function as an adaptive, supplementary strategy that becomes increasingly dominant as value-based estimations grow more difficult.

Despite their adaptive appeal though, frequency-driven decisions often lead to suboptimal performance in laboratory tasks and potentially in real-world contexts (Brevers et al., 2013; Verdejo-Garcia et al., 2006). What remains unclear is how an individual's baseline learning and decision-making capacity shapes susceptibility to these effects in uncertain, complex environments. One possibility is that better learners, while still influenced by reward frequency, are more resilient to frequency-based biases and therefore will rely more consistently on value-based strategies. Another possibility is that, because switching to frequency-based processing can itself be adaptive, better learners are more likely to employ this strategy and thus exhibit stronger frequency effects.

To our knowledge, no study has systematically tested these competing possibilities. The present work addresses this gap using a within-subject design in which participants completed the same RL task twice, with and without a reward frequency manipulation. We aimed to examine whether individuals who performed better during training and in the baseline condition were more or less likely to shift towards frequency-based processing when reward frequency was manipulated, leveraging RL computational modeling. Based on prior research (Hu et al., 2025), we hypothesized that switching to frequency-based processing is adaptive. Thus, we predicted that participants who demonstrated stronger learning performance would be more likely to adopt frequency-based strategies and exhibit stronger frequency effects in the frequency condition.

Methods

Task

This study was approved by the local Institutional Review Board (STUDY2024-1012). The task was adapted from the four-option RL paradigm used in Hu et al. (2025). Participants were presented with four options (A, B, C, D), but on each trial, they selected from a pair of only two options. The expected values (EVs) were ranked as C (0.75) > A (0.65) > B (0.35) > D (0.25). The task consisted of two phases: a training phase and a testing phase.

During the 120-trial training phase, participants repeatedly selected from only two fixed option pairs: AB and CD. On each trial, they compared and chose either between option A and B or between option C and D, with the other two options being unavailable. After each choice, they received feedback showing the number of virtual points earned for that trial, along with a running total of their cumulative points (*Figure 1*). Reward outcomes were randomly drawn from normal distributions centered on each option's EV, with variance approximating binomial distributions. *Table 1* shows the specific reward structure. This level of reward variance has been shown to reliably elicit frequency effects when reward frequency is manipulated (Don et al., 2019; Hu et al., 2025). Within each training pair, option A (0.65) is designed to yield significantly higher rewards than option B (0.35), and option C (0.75) is designed to yield significantly higher rewards than option D (0.25), making A and C the dominant options in their respective pairs. However, the value difference between A and C (i.e., only 0.1) is much smaller, which would allow frequency effects to emerge.

After the training phase, participants proceeded to the testing phase, where they must transfer their knowledge and select from the remaining novel pairings (i.e., AD, BD, CA, CB) without feedback. They were told that the options have remained the same, and they need to make their best selections based on what they learned about each option during the training phase. This phase had 80 trials in total, with 20 trials for each pair type. The two training pairs, AB and CD, did not appear during testing.

We employed a within-subject design that manipulated the frequency of training pairs. Each participant completed the task twice under two different conditions. In the baseline condition, participants chose between A and B on half of the training trials and between C and D on the other half (i.e., 60 AB and 60 CD trials). In the frequency condition, however, AB trials appeared twice as often as CD trials (i.e., 80 AB vs. 40 CD trials), making A more frequently rewarded than C despite its lower EV. We consider CA trials as critical trials because our frequency manipulation creates a direct conflict in the CA pairing, where the slightly less rewarding option, A, is rewarded more frequently than the more rewarding option, C, in the frequency condition. Prior studies have shown that this base-rate manipulation encourages a preference shift from C to A, demonstrating frequency effects (Don et al., 2019; Don & Worthy, 2022; Hu et al., 2025).

Participants were instructed to maximize their cumulative points by learning which options were most rewarding. Choice stimuli were four fractal images randomly drawn from a pool of 12 fractal images, with on-screen positions and image assignments randomized. To avoid any semantic or ordinal priming associated with letters A, B, C, and D, the four options were randomly recoded using arbitrary labels, A, K, L, and S. While option A and B and option C and D always appeared together as pairs, their placement on the screen randomly varied across participants (e.g., ABCD, CDAB, BADC). Trial types were intermixed within each session, and the order of both training and testing trials was randomly shuffled for each participant.

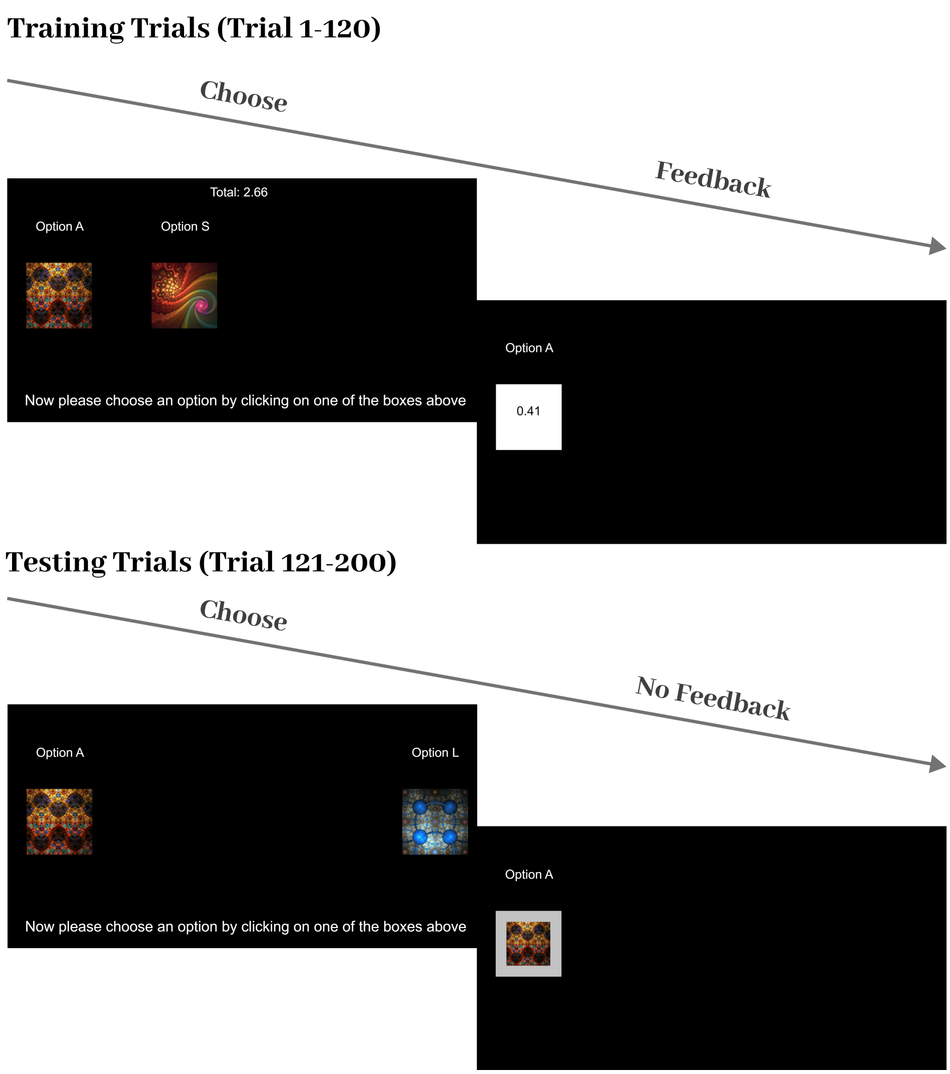
Table 1. Reward Structure.

	Option			
Condition	A	B	C	D

Baseline	N(M,SD)	.65(.48)	.35(.48)	.75(.43)	.25(.43)
	base-rate	1		1	
Frequency	N(M,SD)	.65(.48)	.35(.48)	.75(.43)	.25(.43)
	base-rate	2		1	

Reward Structure. N(M, SD) indicates continuous normal distributions of rewards for each option, where M is the m ean and SD is the standard deviation. “Base-rate” indicates how frequently each choice pair is presented during training, relative to the other choice pair. For example, 2:1 means the first pair (i.e., AB) is presented twice as often as the second pair (i.e., CD). The standard deviations approximated binomial variance, calculated as $SD = (EV_1 \times EV_2)^{0.5}$.

Figure 1. Example Trial Sequence.



Example Trial Sequences. During the first 120 training trials, participants selected from either the two left options or the two right options (i.e., AB or CD). After each choice, they received feedback on the points earned and saw their cumulative total at the top of the screen. In the 80 testing trials, they selected from the remaining four novel pairings without any cumulative point displays or feedback.

Participants

A priori power analysis indicated that 449 participants would provide 95% power to detect a weak-to-moderate effect size in paired samples ($d_z = 0.2$) at an alpha level of .01 (Faul et al., 2007). Based on previous studies in our lab, we anticipated that some participants might show little evidence of learning during training, respond inattentively, or be lost due to technical errors. To account for this, we planned to recruit approximately 500 participants. In total, 501 undergraduate students participated, all of whom provided informed consent and received partial course credit for their participation. Six participants were excluded for inattentive responding, defined as consistently selecting the same option throughout the entire training session for both training pairs (e.g., always choosing A in AB trials and C in CD trials). A final sample of 495 par ticipants went into data analysis.

The mean age was 18.93 years (*SD* = 1.01). The sample consisted of 345 females, 148 males, 1 participant who self-identified as “other”, and 1 who preferred not to answer. Racial distribution was as follows: 370 White, 60 Asian, 31 multiracial, 12 Black or African American, 3 American Indian or Alaska Native, 1 Native Hawaiian or Other Pacific Islander, and 9 who preferred not to answer. With respect to ethnicity, 135 participants identified as Hispanic or Latino, 9 preferred not to answer, and the remainder identified as non-Hispanic.

Procedures

The experiment was administered online. Participants voluntarily enrolled through the u niversity Psychology Department’s research participant pool and received partial course credit for their participation. Upon enrollment, participants were directed to the study via a secure link hosted on the university JATOS server, where the experiment would launch automatically.

The task was developed using the jsPsych JavaScript library for creating online behavioral experiments. Participants completed the RL task twice, once under each condition (i. e., baseline and frequency). After completing the first task, they were directed to fill out a demographic questionnaire and a series of additional surveys included as part of a separate pilot project. They were then returned to complete the second condition of the main task. The entire experiment took approximately one hour to complete, with the two task sessions taking roughly 45-50 minutes in total.

To minimize carryover effects, participants were explicitly instructed that the reward structures were completely different between the two sessions, despite the structural similarity of the tasks, and that prior knowledge should not be transferred between sessions. The order of conditions was randomized and counterbalanced across participants: 252 participants completed the frequency condition first, and 243 completed the baseline condition first. Order effects did not significantly influence behavioral patterns (see *Supplementary Table 1* and *Supplementary Fi gure 1*). After completing both sessions, participants were debriefed, thanked, and informed that they would receive their course credit shortly. The study concluded upon clicking the final submission button.

Data analysis

Behavioral data were analyzed using mixed-effects models implemented via the *lme4* package in *R*. Random intercepts were included at the participant level to account for individual differences given our within-subject design. For analyses involving the percentage of optimal choices, we used the *lmer* function. For trial-level logistic regression models predicting the probability of selecting the optimal option on each trial, we used the *glmer* function.

Computational Models

Beyond traditional behavioral analyses, we employed computational modeling to further disentangle participants' use and weighting of distinct decision-making strategies. We focused on two major classes of reinforcement learning rules: Delta and Decay rules. For each class, we included a basic model and two widely used extensions, yielding six individual models (2×3). To capture the relative contribution of different strategies, we also constructed a seventh hybrid model by combining the best-fitting variant from each class, resulting in a total of 7 formal models.

Delta Model

The basic Delta rule model (Sutton & Barto, 1998) is one of the most widely used value-based RL models. It updates the EV by incorporating the prediction error between the EV from the last trial and the actual reward received in the current trial. EV_{t+1} for option j is defined as:

$$EV_{j,t+1} = EV_{j,t} + \alpha \cdot \left(r_t - EV_{j,t}\right) \cdot I_j \# (1)$$

where I_j is an indicator term set to 1 if option j is chosen on trial t, and 0 otherwise; r_t is the reward value; and α is the recency learning parameter, $\alpha \in (0,1)$, with higher α indicating greater weighting of most recent outcomes. When $\alpha = 0$, EVs remain unchanged regardless of new outcomes; when $\alpha = 1$, EV_j is equivalent to the reward received for option j on its most recent selection. In this model, no memory of previous trial instances is retained, making it mean-centered. Consequently, when prediction errors are minimal, the EV does not substantially change with repeated rewards, rendering the model insensitive to reward frequency.

Delta-Prospect-Valence-Learning (Delta-PVL)

The Delta-PVL model (W. Ahn et al., 2008) extends the basic Delta model by no longer assuming linearity, proportionality, and gain-loss symmetry in the subjective representation of EV, as proposed by the prospect theory (Tversky & Kahneman, 1992). Specifically, this model does not assume veridical processing of EV. Instead, it posits that the magnitude of rewards can be transformed by a non-linear shape parameter γ, and that individuals may apply different weights to gains versus losses via a gain–loss weighting parameter λ. The subjective utility is defined as:

$$u_t = \begin{cases} r_t^{\gamma} & if r_t \geq 0 \\ -\lambda |r_t|^{\gamma} & if r_t < 0 \end{cases} \quad \# (2)$$

where the shape parameter γ ($0 < \gamma < 1$) determines the curvature of the utility function. When $\gamma = 1$, all rewards are processed veridically. As γ approaches 0, reward magnitudes are increasingly discounted, and at $\gamma = 0$, all rewards are treated equivalently (i.e., coded as 1) and the magnitude is completely disregarded. The loss aversion parameter λ ($0 < \lambda < 5$) determines the relative weighting of gains and losses. When $\lambda = 1$, gains and losses contribute equally; λ values below 1 indicate greater sensitivity to gains than losses, whereas λ above 1 indicate greater sensitivity to losses than gains.

The computed utility is then entered into the Delta learning rule to update the EV of the chosen option.

$$EV_{j,t+1} = EV_{j,t} + \alpha \cdot \left(u_t - EV_{j,t}\right) \cdot I_j \#(3)$$

As in equation 1, I_j denotes an indicator for the chosen option and α represents the recency learning parameter.

Delta-Asymmetric

Because all option EVs in our task were positive and outcomes below zero were rare, we included an additional Delta extension model that applies a relative rather than absolute weighting rule for gains and losses. The Delta-Asymmetric model (Niv et al., 2012) assigns separate learning rates to positive and negative prediction errors, defining gains and losses based on whether the received reward exceeds or falls short of the expected value for the chosen option. The EVs are then updated using the standard Delta rule:

$$EV_{j,t+1} = \begin{cases} EV_{j,t} + \alpha^{+} \cdot \delta\left(r_t\right) \cdot I_j & \text{if } \delta\left(r_t\right) > 0, \\ EV_{j,t} + \alpha^{-} \cdot \delta\left(r_t\right) \cdot I_j & \text{if } \delta\left(r_t\right) < 0, \end{cases} \#(4)$$

where $\delta\left(r_t\right)$ represents the prediction error, $\delta\left(r_t\right) = r_t - EV_{j,t}$; α^{+} and α^{-} denote the learning rate for positive and negative prediction errors, respectively, $\alpha \in (0,1)$. This model has been shown to effectively account for risk sensitivity in decision-making (Niv et al., 2012).

Decay

The second major class of reinforcement learning rules is the Decay rule (Erev & Roth, 1998). In contrast to the Delta rule, which represents a recency-weighted average of reward value or subjective utility, the Decay rule assumes that an option's EV increases through repeated selection but gradually decays when the option is not chosen. Formally, the Decay rule updates the EV of option *j* as:

$$EV_{j,t+1} = EV_{j,t} \cdot (1 - A) + r_t \cdot I_j \#(5)$$

where A is the decay parameter akin to α in the Delta rule models, $A \in (0, 1)$. Higher A indicates greater weight assigned to recent outcomes. In this model, the EV for each option gradually decays over time and increases only when a reward for that option is received. As a result, options rewarded more frequently accumulate higher EVs. This mechanism has been shown to capture frequency effects, particularly under conditions where reward frequency may alter participants' perception of EV (Don et al., 2019; Don & Worthy, 2022; Hu et al., 2025).

Decay-PVL

The Decay-PVL model (W.-Y. Ahn et al., 2014) parallels the Delta-PVL model. The conversion from actual rewards to subjective utility follows the same functional form as in Equation 2, but the resulting utility is then integrated using the Decay rule, as follows:

$$EV_{j,t+1} = EV_{j,t} \cdot (1 - A) + u_t \cdot I_j \#(6)$$

Decay-Win

We also fit a relative learning model within the Decay class, adapted from the Prediction-Error Decay model (Don et al., 2022). In this Decay-Win model, EV accumulation depends solely on how often an option yields above-average outcomes. Here, rewards are defined relative to whether the obtained outcome exceeds the running average, AV, updated as:

$$AV_{t+1} = AV_t + \alpha \cdot \left(r_t - AV_t\right) \#(7)$$

and the EV is calculated as:

$$EV_{j,t+1} = EV_{j,t} \cdot (1 - \alpha) + 1 \cdot I_j \#(8)$$

where I_j an indicator term set to 1 if $r_t > AV_t$ and 0 otherwise. In this model, only the valence of the outcome (i.e., whether it is a "win" or not) is used to guide people's choices. EVs decay over time, and increments occur only when rewards surpass the overall average. The model ignores exact reward magnitudes and instead tracks only the number of above-average outcomes associated with each option, thereby providing a clean dissociation between frequency-based and value-based processing.

Hybrid Model

Finally, we selected the best-fitting models from each class to form a hybrid model, consisting of the two relative models: Delta-Asymmetric and Decay-Win (see *Results*). The inclusion of the Decay-Win model as the representative of the Decay rule class completely dissociates value-based processing from frequency-based processing: the Delta component (i.e., Delta-Asymmetric) retains no memory of reward frequency, while the Decay component (i.e., Decay-Win) retains no memory of actual reward values. This hybrid configuration also provided the best model fit among all tested combinations of hybrid models (see *Supplementary Table 2*).

All simple models used a *SoftMax* rule to convert EVs into each model's predicted probability of selecting each j alternative on trial t:

$$P\left|C_{j,t}\right| = \frac{e^{\beta \cdot EV_{j,t}}}{\sum_1^{N(j)} e^{\beta \cdot EV_{j,t}}} \#(9)$$

where $\beta = 3^c - 1$ $(0 \leq c \leq 5)$, and c is a log inverse temperature parameter that determines how consistently the option with a higher EV is selected (Yechiam & Ert, 2007). When $c = 0$, choices are random; as c increases, the option with the highest EV is selected more often. For the hybrid model, a free weighting parameter, w, was applied to the choice probabilities generated by each component process and the final predicted probability is calculated as:

$$P\left|C_j(t)\right| = w_{Delta} \cdot \frac{e^{\beta \cdot EV_{Delta,\ j}(t)}}{\sum_1^{N(j)} e^{\beta \cdot EV_{Delta,j}(t)}} + \left(1 - w_{Delta}\right) \cdot \frac{e^{\beta \cdot EV_{Decay,j}(t)}}{\sum_1^{N(j)} e^{\beta \cdot EV_{Decay,j}(t)}} \#(10)$$

Model Fitting and Evaluation

We used the maximum likelihood (ML) approach for model fitting. The negative log likelihood of the parameter set θ, given observed data y and model M, $L(\hat{\theta}|y,M)$, was minimized using the *minimize* function in the *SciPy* library in Python. To avoid local minima, optimization was repeated 100 times with randomly selected starting points for each parameter. All trials except the first trial in each condition were included for model fitting, and the outcome or utility of the first trial in each condition was used to initialize EV values.

Model comparison relied on the Bayesian Information Criterion (BIC) (Schwarz, 1978). For each participant in each condition, we computed:

$$BIC = -2\ln L\left(\hat{\theta}\middle|y,M\right) + K\ln(N) \# (11)$$

where N is the number of observations. In our study, N equals 200 trials in each condition. To evaluate model evidence, we calculated BIC weights (Wagenmakers & Farrell, 2004):

$$w_i(\mathrm{BIC}) = \frac{\exp -\frac{1}{2}\Delta_i(\mathrm{BIC})}{\sum_{k=1}^{K} \exp -\frac{1}{2}\Delta_k(\mathrm{BIC})} \# (12)$$

where $\Delta_i(\mathrm{BIC}) = \mathrm{BIC}_i - min(\mathrm{BIC})$. Lower BIC values indicate better model fit, whereas higher BIC weights reflect stronger relative support for a given model. We also calculated the Bayes Factor (BF_{10}) using:

$BF_{10,Model1} = \exp\left(\frac{BIC_{Model2} - BIC_{Model1}}{2}\right) \# (13)$ with $\mathrm{BF}_{10} > 3$ generally considered significant, representing a moderate advantage in favor of Model 1 (Wagenmakers, 2007).

In addition, we conducted group-level model comparisons using Variational Bayesian Model Selection (VBMS) (Stephan et al., 2009). VBMS treats each model as a random variable and estimates the parameters of a Dirichlet distribution, which are then used to construct a multinomial distribution describing the probability that each model generated the data of a randomly chosen participant. Posterior Dirichlet parameters, α, represent the estimated frequency with which each model best explains individual participants' data. The posterior multinomial parameter, r_k, gives the probability that data from a randomly chosen participants were generated by mode k. Finally, the exceedance probability, φ_k, quantifies the likelihood that a particular model k is more likely than all competing models to generate group-level data. We used BIC to approximate the log evidence and all three metrics were calculated for model comparisons.

Transparency and openness

All data have been made publicly available at the Open Science Framework (OSF) and can be accessed using the following anonymous link: (https://osf.io/zsk4e/?view_only=49f12e186d164ef699da142fab685573). Materials, analysis code and the code behind model fitting have also been made publicly available at https://osf.io/zsk4e/?view_only=49f12e186d164ef699da142fab685573. This study's design and its analysis were not pre-registered.

Results

Overall Behavioral Results

We first examined participants' performance during the training phase using a mixed-effects logistic regression model, predicting trial-wise choices from condition, trial type, and block. This analysis revealed a main effect of trial type ($\beta = 0.145 \pm 0.039$, $t = 3.680$, $p < .001$), with participants demonstrating higher accuracy on CD trials than AB trials, and a main effect of block ($\beta = 0.039 \pm 0.007$, $t = 5.423$, $p < .001$), with a greater proportion of optimal choices observed over time. There was no main effect of condition ($\beta = 0.056 \pm 0.037$, $t = 1.516$, $p = .130$), suggesting comparable learning effects across conditions. Training performance in AB and CD trials was significantly correlated ($\beta = 0.255 \pm 0.044$, $t = 5.814$, $p < .001$), indicating generally consistent learning performance across the two trained pairs (*Supplementary Figure 2*).

The only significant interaction was between trial type and condition ($\beta = -0.161 \pm 0.057$, $t = -2.807$, $p = .005$). As shown in *Figure 2a*, the performance gap between AB and CD trials was significantly reduced in the frequency condition. This may reflect either poorer learning due to reduced exposure to CD trials or increased exploratory behavior, as participants may perceive CD trials as scarce opportunities worth exploring. All other interactions, including trial type × block ($\beta = 0.008 \pm 0.010$, $t = 0.762$, $p = .446$), condition × block ($\beta = -0.003 \pm 0.009$, $t = -0.281$, $p = .779$), and the three-way interaction ($\beta = 0.009 \pm 0.015$, $t = 0.606$, $p = .544$), were insignificant.

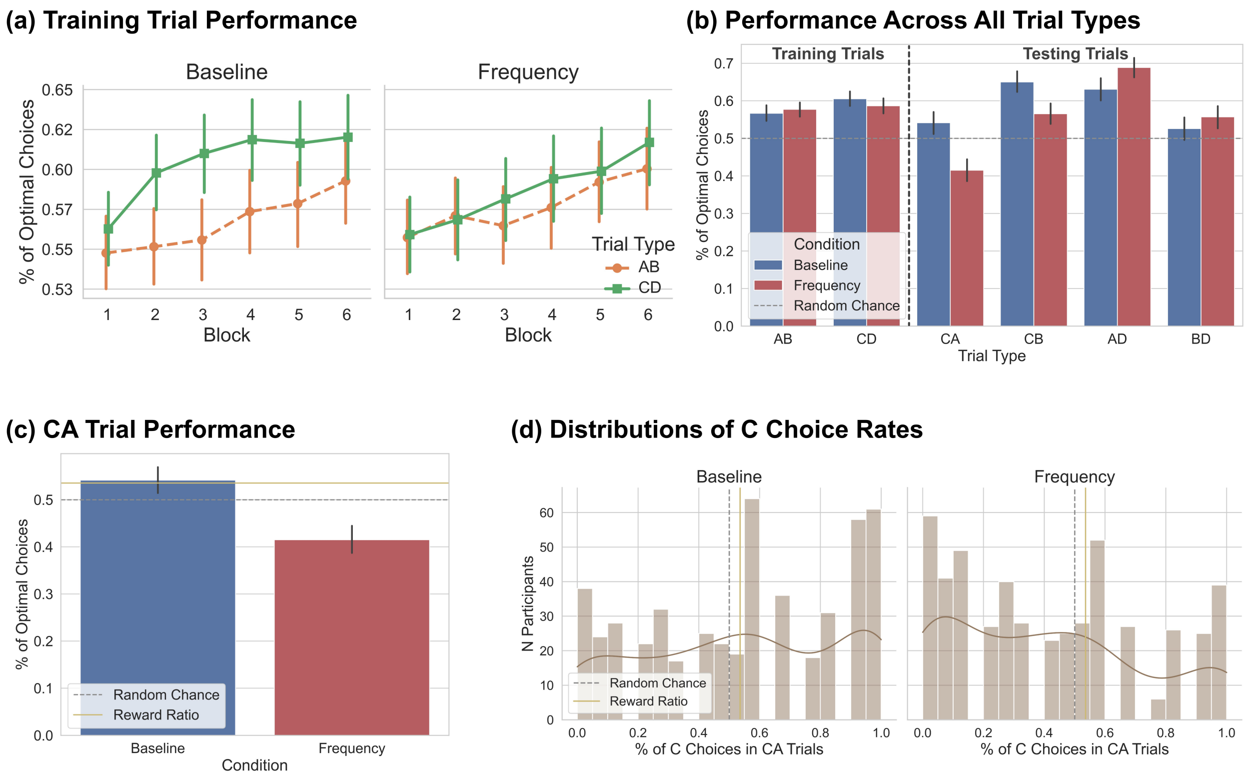
To assess overall performance, we examined the percentage of optimal choices across all six trial types in each condition (*Figure 2b*). For each trial type, we conducted a one-sample *t*-test to compare the observed choice rate against random chance level (0.5) and adjusted for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). After correction, participants selected the optimal option at rates significantly above chance across nearly all trial types across both conditions (see *Supplementary Table 3*), with only two exceptions. In the BD trials under the baseline condition, the proportion of optimal choices ($M = 0.526$) was only numerically above chance, $t(494) = 1.729$, $p_{adjusted} = .084$, and in the CA trials under the frequency condition, participants significantly favored the suboptimal option A over the optimal option C (see details below).

Critical CA Trials

As predicted, in the critical CA trials, participants significantly favored the optimal option C in the baseline condition, $t(494) = 2.824$, $p_{adjusted} = .005$, but showed a reversed preference for the more frequently rewarded yet less valuable option A in the frequency condition, $t(494) = -5.903$, $p_{adjusted} < .001$. This between-condition shift in choice preference was statistically significant based on a paired-sample *t*-test, $t(494) = 6.007$, $p < .001$. To further evaluate this effect, we compared participants' C choice rates against the underlying reward ratio between options C and A, calculated as $\frac{.75}{.75 + .65} \approx .536$. We found that participants' preference for C ($M = .541$, $SD = .329$) closely matched this expected reward ratio in the baseline condition, $t(494) = 0.406$, $p = .685$, whereas in the frequency condition, C choice rates ($M = .415$, $SD = .320$) fell significantly below this ratio, $t(494) = -8.388$, $p < .001$, indicating a robust reward frequency effect. These results aligned with many prior studies using the same paradigm (Don et al., 2019; Don & Worthy, 2022; Hu et al., 2025), which have consistently demonstrated that, under unequal frequencies of reinforcement, people tend to prefer the more frequently rewarded option even when it yields suboptimal outcomes.

Figure 2d shows the distribution of C choice rates across conditions (see *Supplementary Figure 3* for distributions of all trial types). A larger proportion of participants in the frequency condition chose C less often than both random chance and the reward ratio, further confirming the frequency effect.

Figure 2. Behavioral Results



Behavioral Results. (a) Participants generally showed improved performance across blocks. Overall training accuracy did not differ significantly between conditions, but the typical advantage in CD trials over AB trials observed in the baseline condition disappeared in the frequency condition. (b) Participants selected the optimal option at rates significantly above chance across most trial types, with only two exceptions: BD trials in the baseline condition where the optimal choice rates were not significantly different from chance, and CA trials in the frequency condition, where participants significantly favored the more frequently rewarded, yet suboptimal option A. (c) A closer look at CA trials revealed that in the baseline condition, C choice rates closely matched the objective reward ratio, whereas in the frequency condition they fell significantly below both the reward ratio and chance level. (d) The distribution of C choice rates further confirmed this effect and showed that a larger proportion of participants in the frequency condition chose C less often than either the reward ratio or chance level. Error bars indicate 95% CI interval.

Within-Subject Results

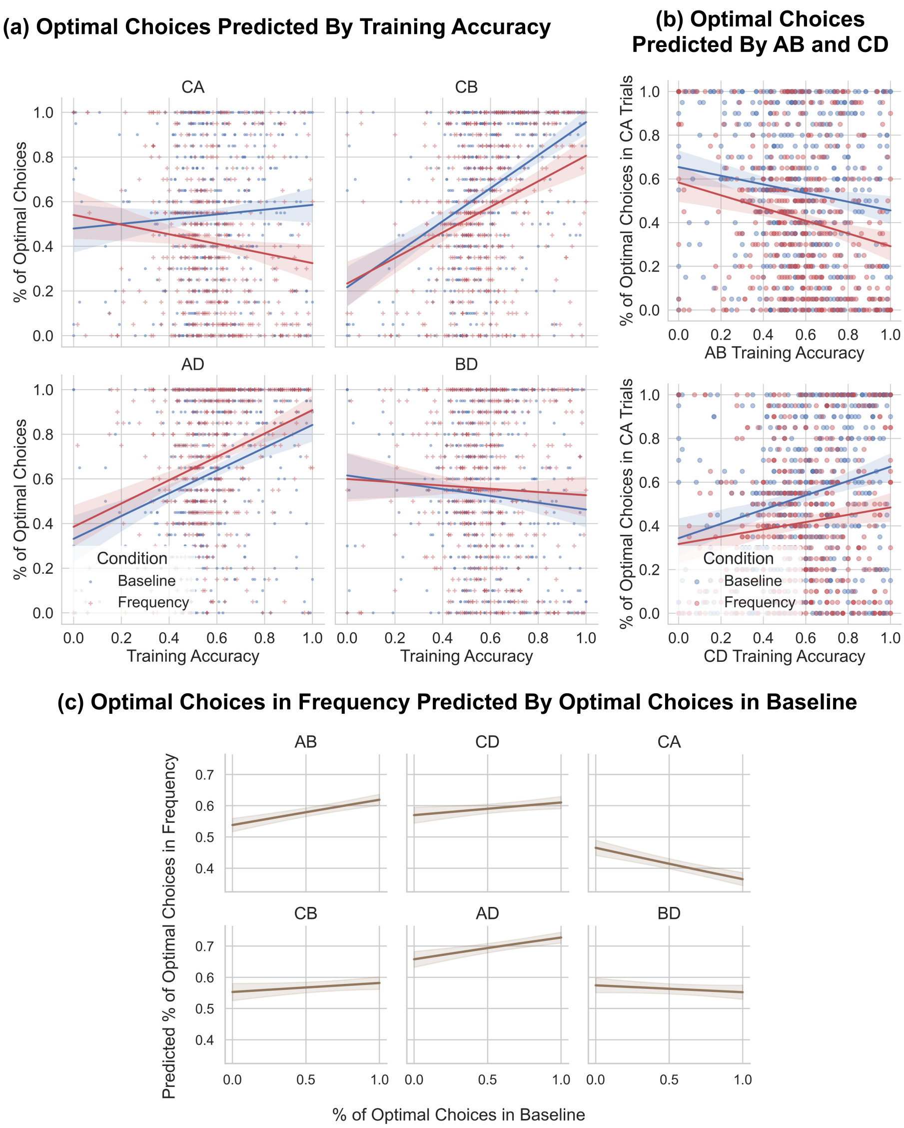
At the core of the present study lies a key question—whether the seemingly irrational, heuristic-driven frequency effect is more pronounced among better learners and decision-makers. We defined "better learners" from two complementary angles: (1) participants who achieved higher overall training accuracy across AB and CD trials, and (2) participants who showed higher C choice rates on CA trials in the baseline condition, indicative of better value-based judgement when reward frequency was not manipulated. As we will show below, these two definitions do not produce divergent results. Importantly, we note that "better learners" here is defined relative to other participants within our sample and does not necessarily imply being a "good learner" in an absolute sense. Using these definitions, we addressed our central question as follows. First, we examined whether participants with higher training accuracy were more likely to select option A in CA trials under the frequency condition. Second, we tested whether participants who demonstrated better value-based learning performance (i.e., higher C choice rates) in the baseline condition were more likely to choose A in CA trials under the frequency condition.

To address the first question, we ran a series of mixed-effects models predicting the percentage of optimal choices during the testing phase from participants' combined training accuracy on AB and CD trials, condition, and their interaction. Within the four testing trial types, higher training accuracy was consistently linked with higher optimal choice rates in AD (β = 0.507 ± 0.072, t = 7.066, p < .001) and CB trials (β = 0.739 ± 0.072, t = 10.287, p < .001), and showed no significant relationship in BD trials (β = -0.153 ± 0.081, t = -1.879, p = .061) regardless of condition. However, in CA trials, we found no main effect of training accuracy (β = 0.106 ± 0.078, t = 1.354, p = .176) but a significant interaction effect (β = -0.321 ± 0.114, t = -2.828, p = .005). In the baseline condition, training accuracy was unrelated—though slightly positively associated—with C choice rates (β = 0.106 ± 0.079, t = 1.333, p = .183), whereas in the frequency condition, the association reversed, with higher training accuracy significantly predicting fewer optimal C choices (β = -0.216 ± 0.081, t = -2.655, p = .008). This reversal was unique to CA trials, as no other trial type showed a change in direction of the training accuracy–optimal choice relationship across conditions (AD: β = 0.017 ± 0.104, t = 0.162, p = .872; CB: β = -0.167 ± 0.105, t = -1.595, p = .111; BD: β = 0.080 ± 0.118, t = 0.677, p = .499; *Figure 3a*). Therefore, we confirmed that while higher training accuracy generally predicted better testing performance, participants with stronger training performance were more likely to choose option A in the frequency condition.

To further disentangle the source of the observed interaction effect, we examined the individual effects of AB and CD training performance on CA trials. We ran two trial-wise logistic regressions predicting accuracy in CA trials by training accuracy in AB or CD trials and condition (see *Supplementary Figure 4* for other trial types). As shown in *Figure 3b*, for AB, higher training accuracy consistently predicted lower optimal choice rates in CA trials (β = -1.073 ± 0.135, t = -7.959, p < .001), indicating that participants who learned AB well (i.e., and therefore chose A frequently) were more biased toward A later. This effect was notably stronger in the frequency condition (β = -0.977± 0.190, t = -5.137, p < .001). For CD, the pattern was reversed. Higher training accuracy in CD trials predicted higher optimal choice rates in CA trials (i.e., favoring C; β = 1.647 ± 0.133, t = 12.413, p < .001), but this effect was attenuated in the frequency condition (β = -1.203 ± 0.185, t = -6.512, p < .001; also see *Supplementary Figure 5*).

Therefore, we decomposed the contributions of the two training pairs to CA performance as follows. First, reward frequency consistently influences testing choices. Better performance on AB trials (i.e., selecting A more often) increases A preference in CA trials, whereas better performance on CD trials (i.e., selecting C more often) increases C preference in CA trials. Critically, however, because AB trials occurred twice as often as CD trials in the frequency condition, participants with overall better learning performance necessarily experienced and reinforced A more often than C. Even without an interaction, this asymmetry alone would bias participants toward choosing A more and produce frequency effects. Second, the significant interaction effects amplify this asymmetry, as participants with higher AB accuracy showed an even steeper A preference in the frequency condition relative to baseline, whereas the positive influence of CD accuracy on C selection was noticeably attenuated. Together, the stronger A and weaker C effects explain why the frequency manipulation disproportionately increases frequency effects among participants with higher training performance.

Figure 3. Within-Subject Results



Within-Subject Results. (a) For CB and AD testing trials, higher training accuracy was consistently associated with higher optimal choice rates in the testing phase. This association was insignificant for BD trials in both conditions. In contrast, CA trials revealed a significant interaction, where training accuracy was positively (insignificant) related to C choice rates in the baseline condition but negatively related in the frequency condition. This suggests that although better training performance generally promoted optimal test-phase choices, it led participants to show stronger frequency effects when reward frequency was manipulated. (b) As we further decouple the unique contributions of AB and CD learning, AB accuracy was consistently associated with greater A choice rates and CD accuracy was consistently associated with greater C choice rates in CA trials. Interaction effects further revealed a steeper AB–CA slope and a flatter CD–CA slope in the frequency condition. This stronger A influence and weaker C influence, coupled with the base-rate imbalance between AB and CD trials in the frequency condition, explains why high-performing participants were numerically more likely to choose C in the baseline condition, but more likely to choose A in the frequency condition (also see *Supplementary Figure 5* for model-based predictions controlling for individual-level variance). (c) Across AB, CD, CB, and AD trials, optimal choice rates were positively correlated between the two conditions, indicating consistent decision-making patterns. For CA trials, however, this association was nonsignificant and numerically negative. Error bars indicate 95% CI interval.

Next, we ran a series of mixed-effects logistic regression models predicting the probability of selecting the optimal option in the frequency condition as a function of the corresponding proportion of optimal choices in the baseline condition, both overall and separately for trial type. Overall, participants' baseline performance significantly predicted their performance in the frequency condition after controlling for trial types ($\beta = 0.061 \pm 0.029$, $t = 2.111$, $p = .035$), suggesting behavioral consistency across conditions. For specific trial types, this positive association held for AB ($\beta = 0.572 \pm 0.258$, $t = 2.213$, $p = .027$), CD ($\beta = 0.606 \pm 0.245$, $t = 2.477$, $p = .013$), AD ($\beta = 0.774 \pm 0.342$, $t = 2.263$, $p = .024$), CB ($\beta = 0.777 \pm 0.329$, $t = 2.357$, $p = .018$) trials, but not for CA ($\beta = -0.412 \pm 0.313$, $t = -1.315$, $p = .188$) or BD trials ($\beta = -0.155 \pm 0.344$, $t = -0.449$, $p = .653$)—both of which involved closely valued options with different reward frequencies during training. Critically, the slope for CA trials was significantly different—and notably negative—compared to AB ($\beta = -0.747 \pm 0.091$, $t = -8.175$, $p < .001$), CD ($\beta = -0.582 \pm 0.098$, $t = -5.945$, $p < .001$), CB ($\beta = -0.532 \pm 0.096$, $t = -5.542$, $p < .001$), BD ($\beta = -0.325 \pm 0.096$, $t = -3.372$, $p < .001$), and AD ($\beta = -0.742 \pm 0.103$, $t = -7.238$, $p < .001$) trials (*Figure 3c*). These findings suggest that while participants' selection of optimal choices was overall consistent across conditions, their choice of C in CA trials was uncorrelated, if not negatively correlated, between baseline and frequency conditions. In other words, participants who strongly preferred the optimal option C under baseline conditions did not maintain that preference when option A was more frequently rewarded, suggesting a strategic shift away from value-based decision-making under frequency manipulation.

Model Fitting Results

Finally, we applied computational modeling to quantify the relative contributions of reward frequency-based and value-based strategies in each condition. As shown in *Table 1*, the Decay-Win and Hybrid models consistently provided the best fits across participants. The Hybrid model yielded the best average fit as indexed by BIC, while the Decay-Win model accounted for the largest number of individual best fits. The overall advantage of Decay-class models over Delta-class models suggests that participants relied on reward frequency during learning in both conditions. In particular, the superior performance of the Decay-Win model, even relative to alternative models within the Decay-class which encode precise numerical reward values (e.g., Decay PVL), indicates that participants may have relied strongly on binary, reward/non-reward tallies when internally updating EV. As option A receives more "wins", or above-average outcomes, in the frequency condition, this dichotomous processing likely reflects a fundamental mechanism through which reward frequency biases arise at the population level. Previous modeling work similarly demonstrates that incorporating binary outcome processing improves model fit and captures reward-frequency-related biases in choice behavior (Hu et al., 2025).

Table 1. Model Fitting Results

	c	$\alpha/A/\alpha$-pos	α-neg	γ	λ	w	BIC_{avg}	BIC_{weight}	BF_{10}	$N_{best\ fit}$	VB α	VB r_k	VB φ_k
Baseline													
Delta	1.708	0.287					252.002	<.001	5316.431	94	54.748	0.109	<.001
Delta PVL	1.769	0.302		0.548	1.862		256.015	<.001	39527.978	12	4.597	0.009	<.001
Delta Asymmetric	1.686	0.352	0.261				242.478	0.018	45.449	43	35.244	0.070	<.001
Decay	0.356	0.211					240.180	0.056	14.404	50	44.905	0.089	<.001
Decay PVL	0.345	0.193		0.357	2.118		242.420	0.018	44.154	30	17.254	0.034	<.001

Decay Win	0.423	0.201					239.151	0.094	8.611	136	**183.276**	**0.365**	**.874**
Hybrid	1.698	0.237	0.365			0.528	**234.845**	**0.813**	-	130	161.976	0.323	.126
Frequency													
Delta	1.766	0.299					253.251	<.001	5220.613	98	69.907	0.139	<.001
Delta PVL	1.808	0.332		0.562	2.001		256.807	<.001	30897.749	20	13.673	0.027	<.001
Delta Asymmetric	1.796	0.382	0.240				245.337	0.008	99.864	34	18.770	0.037	<.001
Decay	0.379	0.225					241.835	0.048	17.336	75	56.601	0.113	<.001
Decay PVL	0.374	0.221		0.340	2.084		244.168	0.015	55.648	21	16.873	0.034	<.001
Decay Win	0.416	0.195					240.334	0.101	8.185	129	**189.481**	**0.377**	**.998**
Hybrid	1.664	0.274	0.381			0.517	**236.130**	**0.828**	-	118	136.695	0.272	.002

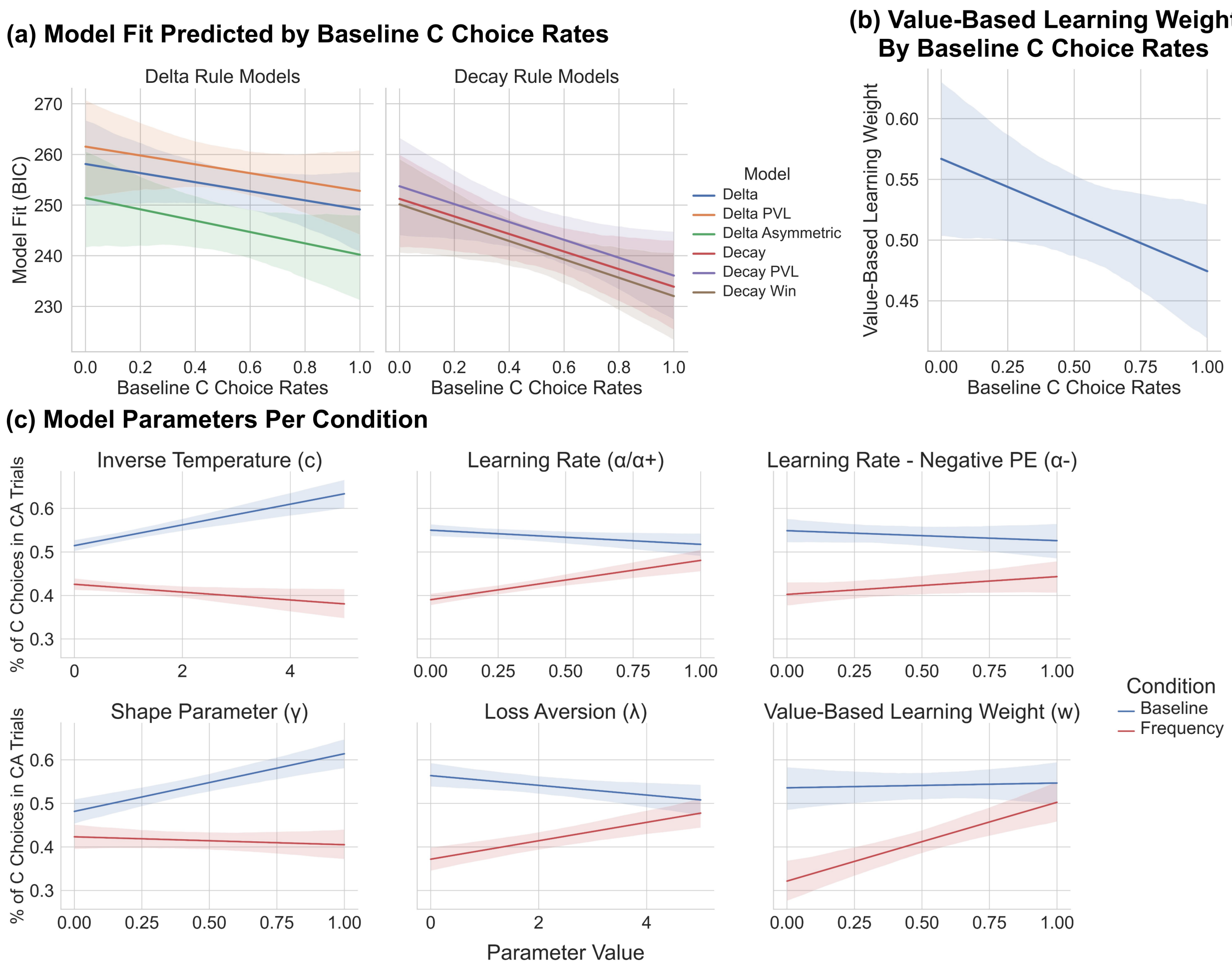
Model Fitting Results. This table summarizes the model fitting results. The parameter c is the inverse log temperature parameter. A higher c means the participant is more likely to stick with the option that has a theoretically higher EV. The parameter of α, A, or *α-pos* (i.e., α for positive prediction errors in asymmetric learning models) capture recency effects, reflecting the extent to which recent outcomes influence subsequent choices; higher values indicate faster decay of past experiences and stronger reliance on recent samples. The parameter *a-neg* in the asymmetric models represents the learning rate for negative prediction errors. The shape parameter γ determines the curvature of subjective utility as a function of reward magnitude, with lower values reflecting stronger discounting of reward magnitude. The loss-aversion parameter λ weights losses relative to gains, with higher values reflecting greater loss aversion. The weighting parameter w in the hybrid model specifies the relative contribution of value-based, Delta-rule processing, such that lower values reflect greater reliance on frequency-based processing. Model fit was evaluated using BIC, BIC weights, Bayes Factors (BF_{10}) and Variational Bayesian Model Selection (VBMS). Lower BIC values indicate better model fit, while higher BIC weights indicate stronger relative support. BIC-weights sum to one. BF_{10} represents the Bayes Factor difference between a given model and the best-fitting model. Conventionally, a BIC difference of 0–2 provides little support for the better model, 4–7 indicates moderate support, and differences of 10 or more indicate strong support. A BF_{10} greater than 3 is considered significant evidence. Finally, the last three columns of the table report VBMS results, with BIC approximating log evidence: VB α values represent estimated model frequencies; VB r_k values indicate the probability that model k generated the data for a randomly selected participant; and VB φ_k reflects the exceedance probability that model k is more likely than all alternatives at the group level.

Parameter Analysis

To further examine within-subject changes in strategy use and identify individuals who exhibit stronger frequency effects, we analyzed whether participants' baseline behavior could predict reliance on frequency-based strategies in the frequency condition. We first focused on simpler, single-process models. A mixed-effects model was conducted to predict model fit (i.e., BIC) of the three Decay rule models in the frequency condition with participants' C choice rates in the baseline condition. We controlled for model type to reduce model-specific noise. Results revealed that higher C choice rates in the baseline condition significantly predicted better fit of D ecay rule models in the frequency condition ($\beta = -17.699 \pm 6.939$, $t = -2.551$, $p = .011$), suggesting that individuals who learned value-based contingencies well in baseline condition were more likely to be better captured by frequency-sensitive Decay rule models when reward frequency was manipulated (*Figure 4a*). Crucially, this relationship did not hold for frequency-in sensitive Delta rule models ($\beta = -9.632 \pm 6.353$, $t = -1.516$, $p = .130$), whose fit was unrelated to participants' baseline C choice rates. We found no evidence for the usage of alternative heuristic processes beyond reward-frequency-based biases across our comparison models (*Supplementary Figure 6*).

Next, we examined the hybrid model to assess whether participants who selected option C more frequently in the baseline condition were more likely to shift toward a frequency-based strategy under frequency manipulation. Specifically, we ran a general linear model predicting the model-inferred Delta weights (i.e., the relative reliance on value-based learning) in the frequency condition based on participants' C choice rates in the baseline condition, controlling for training performance. The results revealed a significant negative correlation: participants who chose C more often in the baseline condition showed lower Delta weights in the frequency condition (β = -0.103 ± 0.052, t = -1.977, p = .049), indicating greater reliance on the frequency-based process (*Figure 4b*). Importantly, this pattern was consistent across alternative, less well-fitting hybrid configurations where different Delta and Decay variants were combined (*Supplementary Figure 7*). Together, these findings suggest that individuals who previously engaged in more rational, value-based learning were particularly susceptible to switching to a heuristic, frequency-driven strategy when reward frequencies were unequal and demonstrating stronger frequency effects.

Figure 4. Model Fitting Results.



Model Fitting Results. (a) Higher C choice rates in the baseline condition significantly predicted better fit for Decay-class models in the frequency condition, but not for Delta-class models (see *Supplementary Figure 6* for baseline model fits). (b) Examination of the weight parameter in the Hybrid model showed that higher baseline C choice rates predicted reduced reliance on value-based Delta processing in the frequency condition. This suggests that individuals with higher baseline accuracy in CA trials may absorb less information from value-based strategies and rely more on frequency-based processing when reward frequency was manipulated. (c) All model parameters showed significant interaction effects across the two conditions, such that parameters positively associated with C choice rates in the baseline condition predicted the opposite or null effect in the frequency condition, and vice versa. Given the strong parameter correlations across conditions, this pattern suggests that traits leading individuals to favor C in the baseline condition may also predispose them to shift toward A in the frequency condition. Error bars indicate 95% CI interval.

Finally, we looked into the best-fitting model parameters. Parameter estimates were strongly correlated across conditions (β = 0.158 ± 0.010, t = 15.935, p < .001) after controlling for parameter type and model type, indicating high consistency in general decision-making characteristics across conditions. However, mixed-effects models predicting the percentage of C choices in CA trials based on condition and specific parameters revealed significant interactions across all parameters (c: β = -0.031 ± 0.004, t = -7.451, p < .001; α: β = 0.099 ± 0.018, t = 5.560, p < .001; α-*neg*: β = 0.075 ± 0.035, t = 2.128, p = .034; γ: β = -0.136 ± 0.031, t = -4.432, p < .001; λ: β = 0.030 ± 0.006, t = 5.230, p < .001; w: β = 0.170 ± 0.054, t = 3.164, p = .002). Specifically, parameters associated with higher C choice rates in the baseline condition were associated with lower C choice rates in the frequency condition, and vice versa (*Figure 4c*). Given the overall stability of individuals' best-fitting parameters across conditions, these results indicate that individuals whose traits inclined them to favor C in the baseline condition are more likely to shift towards A in the frequency condition.

Discussion

In the science of decision-making, it has long been argued that the use of heuristic, fast-and-frugal strategies may be adaptive (Gigerenzer, 1996; Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 1996; Payne et al., 1988). Yet no study has systematically examined whether such heuristics are more or less likely to emerge in individuals who demonstrate stronger learning performance. In the present study, we addressed this question using a within-subject design in which each participant completed the task twice—once under baseline conditions and once with a reward-frequency manipulation in an environment known to elicit frequency effects (Don et al., 2019; Hu et al., 2025).

Our manipulation successfully altered participants' behavior. Participants strongly preferred the more valuable option C in the baseline condition but shifted to favor the less valuable option A when A was presented twice as often during training in the frequency condition. Within-subject analyses further revealed two key patterns. First, participants with better training performance generally achieved higher accuracy in the testing phase, except in CA trials under the frequency condition, where higher training accuracy predicted a stronger tendency to favor the frequently rewarded but less valuable option A. Second, accuracy across trial types was generally consistent between conditions, whereas in the critical CA trials, C choice rates in the baseline condition did not predict, and even negatively trended with, C choice rates in the frequency condition.

Computational modeling validated these behavioral findings. Participants with higher baseline C choice rates were found to be better fit by frequency-sensitive Decay class models in the frequency condition but not frequency-insensitive Delta class models. In the Hybrid model, which combines a purely value-based and a purely frequency-based process with a weighting parameter, these good learners showed greater reliance on the frequency-based process when the less valuable option was rewarded more often. Together, these results suggest that individuals who demonstrate stronger baseline learning are also those more likely to shift toward frequency-based processing when strong frequency cues are present, thereby exhibiting stronger frequency effects.

Since the IGT gained popularity, it has often been assumed that focusing on reward frequency while neglecting underlying reward values reflects impaired value learning associated with neuropsychological deficits (Bechara, 2000; Bechara et al., 1994, 1996, 1999). However, more recent studies consistently demonstrate that even healthy participants prefer options yielding more frequent rewards, despite their lower average value (Chiu et al., 2008; Don et al., 2019; Horstmann et al., 2012; Hu et al., 2025; Kumar et al., 2019; Lin et al., 2007; Steingroever et al., 2013). This suggests that reward frequency is not merely a marker of deficit, but potentially a fundamental component of human value learning and decision-making. Evidence from both the IGT and its inverted version further underscores this point by demonstrating that frequent small wins outweigh infrequent large losses, leading participants to view the frequently rewarded option as superior (Horstmann et al., 2012; Overman et al., 2004; Steingroever et al., 2013). Conversely, frequent small losses overshadow infrequent large wins, prompting participants to avoid these options even when the long-term value is positive (Lin et al., 2012). In the present study, we extend this literature by showing that reward frequency plays an indispensable role in shaping subjectively perceived values. Crucially, individuals with stronger baseline learning and decision-making performance exhibited more pronounced, seemingly irrational, frequency effects. This pattern supports the view that frequency effects are not simply a sign of flawed learning in complex environments. Instead, they may reflect an adaptive strategy shift, where good learners are able to actively perceive, process, and integrate frequency cues into their decision-making as part of a flexible response to environmental demands.

That said, our findings do not preclude the possibility that a rigid, excessive focus on reward frequency can be maladaptive. Prior research shows that frequency effects at the group level typically emerge only under conditions of highly elevated environmental uncertainty (Hu et al., 2025). Attending exclusively to immediate rewards, without regard to long-term payoffs, in environments that do not necessitate such a strategy switch may signal deficient cognitive functioning, such as low self-control (Pang et al., 2015). Thus, the adaptiveness of seemingly irrational heuristics, such as frequency effects, depends on whether the environment calls for a flexible shift away from purely value-based decision-making, but may become harmful when value-based strategies remain tractable. Overall, our findings emphasize that frequency effects are neither a mere cognitive flaw nor a one-size-fits-all strategy. Rather, it may serve as a context-dependent adaptive tool that enables decision-makers to flexibly balance efficiency and accuracy when navigating uncertain environments.

Limitations

One limitation of the present study is that the adaptive switching from value-based decision-making to seemingly irrational mental shortcuts was examined only in the context of frequency effects in reinforcement learning. Other heuristics that are not necessarily tied to reward frequency, such as take-the-best, one-clever-clue, and fast-and-frugal trees (Gigerenzer & Gaissmaier, 2011), require further investigation to determine whether they, too, emerge more strongly in better learners. Relatedly, although we found no evidence that better learners were picking up alternative heuristic processes beyond reward frequency in our comparison model set (e.g., gain-loss asymmetry assumed by PVL models, or positive-negative prediction error asymmetry assumed by the Delta Asymmetric model), it remains possible that they may engage other heuristics uncaptured by the set of comparison models used here. In addition, it remains unclear whether such strategy switching persists in environments characterized by low to moderate uncertainty. We speculate that individuals with stronger learning performance may be more resilient to frequency-based biases when the environment is relatively simple to grasp, but this possibility awaits empirical testing. Finally, this study was not preregistered. Future preregistered replication work that extends our findings or tests our predictions will be critical for further validating and refining our results.

Conclusions

In conclusion, the present study demonstrates that frequency effects, as long considered markers of impaired value learning, may instead reflect an adaptive shift in strategy, particularly among individuals with stronger learning performance. By combining behavioral analyses with computational modeling, we showed that better learners not only displayed stronger frequency effects under reward-frequency manipulation but were also best captured by frequency-sensitive models. These findings suggest that what appears to be irrational bias may, under conditions of uncertainty, represent a flexible adaptation when value-based strategies are costly or unreliable. Ultimately, clarifying when and why individuals adopt such seemingly irrational biases will advance our understanding of the balance between rational value learning and adaptive shortcuts in human decision-making, offering deeper insight into the nature of human rationality in ecological, real-world contexts.

References

Ahn, W., Busemeyer, J. R., Wagenmakers, E., & Stout, J. C. (2008). Comparison of Decision Learning Models Using the Generalization Criterion Method. *Cognitive Science*, *32*(8), 1376–1402. https://doi.org/10.1080/03640210802352992

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: evidence from computational modeling with pure users. *Frontiers in Psychology*, *5*, 849.

Bechara, A. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex*, *10*(3), 295–307. https://doi.org/10.1093/cercor/10.3.295

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making. *The Journal of Neuroscience*, *19*(13), 5473–5481. https://doi.org/10.1523/JNEUROSCI.19-13-05473.1999

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex. *Cerebral Cortex*, *6*(2), 215–225. https://doi.org/10.1093/cercor/6.2.215

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brevers, D., Bechara, A., Cleeremans, A., & Noël, X. (2013). Iowa Gambling Task (IGT): twenty years after – gambling disorder and IGT. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00665

Chiu, Y.-C., & Lin, C.-H. (2007). Is deck C an advantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions*, *3*(1), 37. https://doi.org/10.1186/1744-9081-3-37

Chiu, Y.-C., Lin, C.-H., Huang, J.-T., Lin, S., Lee, P.-L., & Hsieh, J.-C. (2008). Immediate gain is long-term loss: Are there foresighted decision makers in the Iowa Gambling Task? *Behavioral and Brain Functions*, *4*(1), 13. https://doi.org/10.1186/1744-9081-4-13

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73. https://doi.org/https://doi.org/10.1016/0010-0277(95)00664-8

Don, H. J., Davis, T., Ray, K. L., McMahon, M. C., Cornwall, A. C., Schnyer, D. M., & Worthy, D. A. (2022). Neural regions associated with gain-loss frequency and average reward in older and younger adults. *Neurobiology of Aging*, *109*, 247–258. https://doi.org/10.1016/j.neurobiolaging.2021.10.001

Don, H. J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*. https://doi.org/10.1016/j.cognition.2019.104042

Don, H. J., & Worthy, D. A. (2022). Frequency effects in action versus value learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1311–1327. https://doi.org/10.1037/xlm0000896

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 848–881.

Fang, X., Singh, S., & Ahluwalia, R. (2007). An Examination of Different Explanations for the Mere Exposure Effect. *Journal of Consumer Research*, *34*(1), 97–103. https://doi.org/10.1086/513050

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gigerenzer, G. (1996). The Psychology of Good Judgment. *Medical Decision Making*, *16*(3), 273–280. https://doi.org/10.1177/0272989X9601600312

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733–758. https://doi.org/10.1037/0033-295X.114.3.733

Horstmann, A., Villringer, A., & Neumann, J. (2012). Iowa Gambling Task: There is more to consider than long-term outcome. Using a linear equation model to disentangle the impact of outcome and frequency of gains and losses. *Frontiers in Neuroscience*, *6*, 61.

Hu, M., Don, H. J., & Worthy, D. A. (2025). Distributional dual-process model predicts strategic shifts in decision-making under uncertainty. *Communications Psychology*, *3*(1), 61. https://doi.org/10.1038/s44271-025-00249-y

Kim, Y.-T., Sohn, H., & Jeong, J. (2011). Delayed transition from ambiguous to risky decision making in alcohol dependence during Iowa Gambling Task. *Psychiatry Research*, *190*(2–3), 297–303. https://doi.org/10.1016/j.psychres.2011.05.003

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychological Science*, *28*(9), 1321–1333. https://doi.org/10.1177/0956797617708288

Kumar, R., Janakiprasad Kumar, K., & Benegal, V. (2019). Underlying decision making processes on Iowa Gambling Task. *Asian Journal of Psychiatry*, *39*, 63–69. https://doi.org/10.1016/j.ajp.2018.12.006

Lin, C.-H., Chiu, Y.-C., & Huang, J.-T. (2009). Gain-loss frequency and final outcome in the Soochow Gambling Task: A reassessment. *Behavioral and Brain Functions*, *5*, 1–9.

Lin, C.-H., Chiu, Y.-C., Lee, P.-L., & Hsieh, J.-C. (2007). Is deck B a disadvantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions*, *3*(1), 16. https://doi.org/10.1186/1744-9081-3-16

Lin, C.-H., Song, T.-J., Chen, Y.-Y., Lee, W.-K., & Chiu, Y.-C. (2013). Reexamining the Validity and Reliability of the Clinical Version of the Iowa Gambling Task: Evidence from a Normal Subject Group. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00220

Lin, C.-H., Song, T.-J., Lin, Y.-K., & Chiu, Y.-C. (2012). Mirrored Prominent Deck B Phenomenon: Frequent Small Losses Override Infrequent Large Gains in the Inverted Iowa Gambling Task. *PLoS ONE*, *7*(10), e47202. https://doi.org/10.1371/journal.pone.0047202

McCammon, I., & Hägeli, P. (2007). An evaluation of rule-based decision tools for travel in avalanche terrain. *Cold Regions Science and Technology*, *47*(1–2), 193–206. https://doi.org/10.1016/j.coldregions.2006.08.007

Niv, Y., Edlund, J. A., Dayan, P., & O’Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *The Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

North, N. T., & O’Carroll, R. E. (2001). Decision making in patients with spinal cord damage: afferent feedback and the somatic marker hypothesis. *Neuropsychologia*, *39*(5), 521–524. https://doi.org/https://doi.org/10.1016/S0028-3932(00)00107-X

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition*, *37*(5), 632–643. https://doi.org/10.3758/MC.37.5.632

Overman, W. H., Frassrand, K., Ansel, S., Trawalter, S., Bies, B., & Redmond, A. (2004). Performance on the IOWA card task by adolescents and adults. *Neuropsychologia*, *42*(13), 1838–1851. https://doi.org/10.1016/j.neuropsychologia.2004.03.014

Pang, B., Otto, A. R., & Worthy, D. A. (2015). Self-Control Moderates Decision-Making Behavior When Minimizing Losses versus Maximizing Gains. *Journal of Behavioral Decision Making*, *28*(2), 176–187. https://doi.org/10.1002/bdm.1830

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*(2), 258–276. https://doi.org/10.1016/j.actpsy.2007.05.004

Ritter, L. M., Meador-Woodruff, J. H., & Dalack, G. W. (2004). Neurocognitive measures of prefrontal cortical dysfunction in schizophrenia. *Schizophrenia Research*, *68*(1), 65–73. https://doi.org/https://doi.org/10.1016/S0920-9964(03)00086-0

Sallum, I., Mata, F., Miranda, D. M., & Malloy-Diniz, L. F. (2013). Staying and shifting patterns across IGT trials distinguish children with externalizing disorders from controls. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00899

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*(3), 610–628. https://doi.org/10.1037/0033-295X.112.3.610

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Shurman, B., Horan, W. P., & Nuechterlein, K. H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. *Schizophrenia Research*, *72*(2–3), 215–224. https://doi.org/10.1016/j.schres.2004.03.020

Steingroever, H., Wetzels, R., Horstmann, A., Neumann, J., & Wagenmakers, E.-J. (2013). Performance of healthy participants on the Iowa Gambling Task. *Psychological Assessment*, *25*(1), 180–193. https://doi.org/10.1037/a0029929

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review*, *11*(4), 742–747. https://doi.org/10.3758/BF03196629

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Toplak, M. E., Jain, U., & Tannock, R. (2005). Executive and motivational processes in adolescents with Attention-Deficit-Hyperactivity Disorder (ADHD). *Behavioral and Brain Functions*, *1*(1), 8. https://doi.org/10.1186/1744-9081-1-8

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Upton, D. J., Kerestes, R., & Stout, J. C. (2012). Comparing the Iowa and Soochow Gambling Tasks in Opiate Users. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00034

Verdejo-Garcia, A., Bechara, A., Recknor, E. C., & Perez-Garcia, M. (2006). Decision-Making and the Iowa Gambling Task: Ecological validity in individuals with substance dependence. *Psychologica Belgica*, *46*(1–2), 55. https://doi.org/10.5334/pb-46-1-2-55

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196.

Wilder, K. E., Weinberger, D. R., & Goldberg, T. E. (1998). Operant conditioning and the orbitofrontal cortex in schizophrenic patients: unexpected evidence for intact functioning. *Schizophrenia Research*, *30*(2), 169–174. https://doi.org/10.1016/S0920-9964(97)00135-7

Wu, C. M., Schulz, E., Pleskac, T. J., & Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific Reports*, *12*(1), 4122. https://doi.org/10.1038/s41598-022-07901-1

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*(3), 387–402. https://doi.org/10.3758/BF03193783

Yechiam, E., & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology*, *51*(2), 75–84. https://doi.org/10.1016/j.jmp.2006.11.002

Zajonc, R. B., Markus, H., & Wilson, W. R. (1974). Exposure effects and associative learning. *Journal of Experimental Social Psychology*, *10*(3), 248–263. https://doi.org/10.1016/0022-1031(74)90071-7