Can Visual Fixations Explain Context-Dependent Reinforcement Learning?

Melanie J. Touchard & William M. Hayes
Department of Psychology, Binghamton University, NY, USA

Abstract

Context-dependent reinforcement learning (RL) challenges the assumption that decision makers encode the absolute values of choice outcomes. This study investigates whether the associated choice biases arise from a relative encoding of outcomes or an alternative mechanism involving cumulative reward learning and selective attention to outcomes. Using eye tracking, participants completed a RL task where choice options were initially learned in fixed contexts before being tested in novel pairings. Results revealed an overall preference for options that were contextually favored in the learning phase, even when these preferences violated expected value maximization. Computational model comparisons demonstrated that hybrid encoding models, incorporating absolute and relative values, provided the best overall account of individual behavior. While eye fixations on choice outcomes decreased over trials, fixation-dependent RL models did not fit the data well, suggesting that overt visual attention patterns do not fully explain context-dependent choice biases.

Keywords: value encoding; computational modeling; eye tracking; decision making

Introduction

Decision making often requires individuals to select options that yield the highest possible reward. Normative theories suggest that individuals accomplish this by comparing the subjective values, or expected utilities, of the various options (Savage, 1954; Von Neumann & Morgenstern, 1947). In these theories the subjective value of an option is assumed to be unaffected by the surrounding context of other options.

However, evidence from several studies indicates that subjective valuation is highly context-dependent: The same option can seem attractive or unattractive depending on the other options that are available (Hunter & Daw, 2021; Palminteri & Lebreton, 2021; Tversky & Simonson, 1993). This is especially true when the same options are repeatedly encountered in specific contexts (Bavard et al., 2021; Hayes & Wedell, 2023a; Pompilio & Kacelnik, 2010). Consider a choice environment with four alternatives that have different expected values (EVs). If Option A (EV = 15) is always encountered with Option B (EV = 18), individuals may simply learn that B is "good" and A is "bad," without learning the absolute expected rewards. Similarly, if Option C (EV = 21) is always encountered with Option D (EV = 24), individuals may learn that D is good, and C is bad. This kind of relative valuation can lead to successful reward maximization when the learning contexts are fixed, but it can lead to suboptimal choices when the options are presented in novel configurations. For example, if Option B were later paired with Option C, an individual who only learned the relative values would choose B, even though C gives more average reward. Many studies have shown that people do indeed exhibit these kinds of suboptimal choice biases in laboratory settings (e.g., Bavard et al., 2018; 2021; Hayes & Wedell, 2023a; Klein et al., 2017; Molinaro & Collins, 2023; Palminteri et al., 2015).

Computational modeling can be used to understand how individuals dynamically learn and adjust their choices based on feedback from the environment. To understand how context shapes this process, the dominant approach has been to augment standard reinforcement learning (RL) models with some type of valuation mechanism that incorporates relative comparisons among outcomes (e.g., Bavard & Palminteri, 2023; Hayes & Wedell, 2023b). The general idea is that the subjective value of an outcome depends on how it compares to other outcomes from the same local context. However, there is another potential mechanism that can explain context-dependent choice biases in RL, and which does not require a comparative valuation process. As we show in the Results, models that assume absolute, context-independent valuation can produce similar choice biases if they track cumulative reward instead of average reward (Don et al., 2019), and if the outcomes from chosen options are attended to and encoded more frequently than the outcomes from unchosen options (Ashby & Rakow, 2016; Bault et al., 2016). In short, this occurs because options that are contextually advantaged will tend to be chosen more often, and thus more of their outcomes will end up being encoded, increasing the (subjective) cumulative rewards for these options relative to other options. We contend that this plausible alternative explanation, overlooked by prior studies of context-dependent RL, can be tested using a combination of eye-tracking and model comparison.

Both cumulative reward learning and an attentional advantage for obtained over foregone outcomes are supported by prior work. Several studies have demonstrated that cumulative reward (i.e., decay rule) models outperform average reward (i.e., delta rule) models in other RL paradigms (Don et al., 2019; Don & Worthy, 2022; Yechiam & Busemeyer, 2005), yet the models proposed to account for context-dependent RL have all relied on the delta rule (Rescorla & Wagner, 1972). Regarding attentional allocation, Ashby and Rakow (2016) found that outcome fixations in a repeated decision-making task diminished over trials as individuals learned from their choices, and that obtained outcomes were fixated more frequently than foregone outcomes. Other studies have shown that the chosen option's outcome is typically fixated first (Lefebvre et al., 2024) and for a longer duration than

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

the foregone outcome (Bault et al., 2016). Existing models of context-dependent RL do not account for how attention modulates learning by shaping which outcomes are encoded. Yet, other studies have demonstrated that experience-based decisions are well-described by models that accumulate fixation-sampled outcomes, suggesting a tight link between overt visual attention to outcomes and the weight they receive in subsequent choice (Glöckner et al., 2012).

The present study leveraged computational modeling and eye-tracking to test alternative accounts of context-dependent choice in RL. Our primary goal was to test whether context-dependent choice biases are better explained by models that assume a relative encoding of outcomes (e.g., Bavard & Palminteri, 2023; Hayes & Wedell, 2023b), or by models that assume absolute encoding but use a cumulative learning rule and allow for an uneven allocation of attention to obtained and foregone outcomes. Eye fixations during feedback presentation were used as a measure of overt visual attention to outcomes. The following hypotheses were tested:

H1: There will be an aggregate choice preference for options that were relatively favored in their original learning context, even when choosing those options violates expected value maximization (e.g., choosing Option B over Option C in the example above).

H2: The probability of fixating on choice outcomes will decrease across the learning phase, and will be greater overall for the outcomes of chosen options (Ashby & Rakow, 2016; Bault et al., 2016).

H3: Models that assume relative outcome encoding will predict choice better than models that accumulate fixation-sampled absolute outcomes (Glöckner et al., 2012).

Method

Participants

Participants were 50 undergraduate students who participated in exchange for partial course credit (39 women, 12 men; ages 18-27, M_{age} = 19.1). The study was approved by the IRB at the authors' institution.

Apparatus

Gaze position was recorded from the right eye with a table-mounted EyeLink 1000 Plus eye-tracker (sampling rate: 250 Hz). The stimuli were presented with a constant gray background on a 24-inch HD LCD monitor at a resolution of 1920 x 1080 pixels. Participants sat approximately 100 cm from the screen with their heads stabilized in a chin rest. The average validation error across participants was 0.27 degrees of visual angle.

Stimuli and Task

The task was adapted from Hayes and Wedell (2023a). Eight options, associated with Gaussian reward distributions, were randomly assigned to symbols at the start of the session. The options were paired into four fixed learning contexts during the learning phase (Figure 1). Each learning trial involved choosing between the two symbols from a randomly selected context, followed by full feedback from both options, with the outcomes displayed as points. Each symbol pair was presented 30 times in random order for 120 total learning trials. The learning phase was followed by a transfer test that included choices between all 28 possible symbol pairs repeated four times each for 112 total test trials. No feedback was presented during the transfer test. The symbols were presented an equal number of times on the left and right side in both phases.

Procedure

At the start of each session, the system was calibrated to the participant's eye using a 9-point calibration. Participants then read instructions for the learning phase. They were informed that the first part of the task involved choosing between symbol pairs and viewing the points awarded, with the goal of learning which symbols were the most valuable. The instructions stated that in the second part of the task, they would be able to make choices based on what they learned in the first part, and that they would win candy based on the number of correct choices made. No additional details about the transfer test were provided.

Each learning trial began with a fixation to align the participant's gaze in the center of the screen, followed by two symbols presented side-by-side. Participants chose a symbol using the left/right arrow keys. After their choice, participants had to look at a central fixation cross for 300 ms before the outcomes for the chosen and unchosen options were revealed below the symbols with the chosen symbol highlighted. Participants could view the outcomes as long as they preferred before pressing the up arrow key to proceed. Trials were presented in a random order for each participant.

Following the learning phase, participants were invited to take a short break before recalibrating their eye to start the transfer test. Participants were instructed to choose the symbol on each trial that they thought would give more points. Each transfer trial began with a fixation, followed by two symbols presented side-by-side. After choosing a symbol using the arrow keys, the chosen symbol was highlighted for 0.5 seconds. Trial order was randomized. After the choice task, participants completed a memory-based value estimation task on a different computer (results not presented here).

Data Processing

Gaze coordinates were processed using Data Viewer software (EyeLink Data Viewer [version 5.15]), which automatically parsed the raw samples into fixations and saccades. Areas of interest (AOIs) were defined around the option symbols (340 x 340 pixels) and around the outcomes during the learning phase (340 x 100 pixels) (Figure 1). For the present analysis, we focus on fixations occurring during the feedback viewing periods (learning

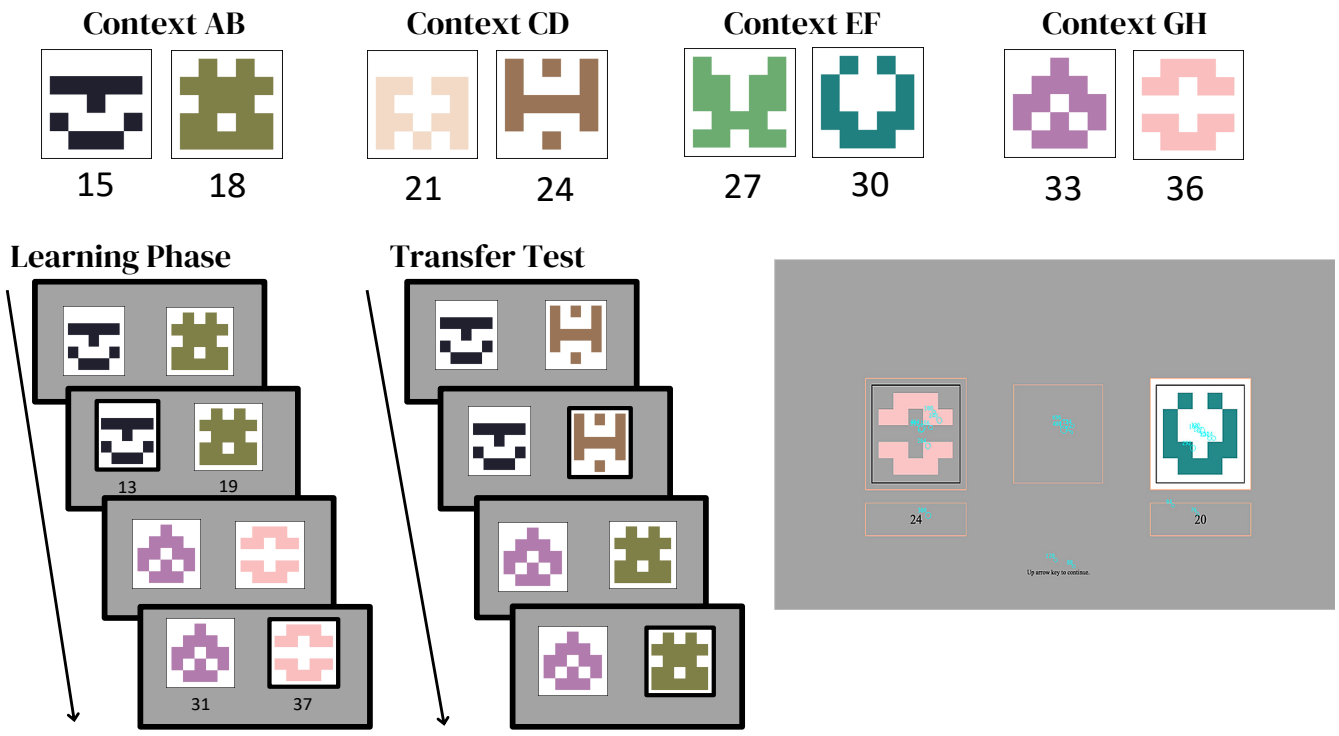

Figure 1: (Top) The four learning contexts (symbol pairs) and the option expected values (EVs). Symbols were randomly assigned to options for each participant. Each option's outcomes were drawn from a Gaussian distribution with the given EV (SD = 2) and rounded to the nearest integer. (Bottom Left) Learning and transfer trials (fixation crosses not shown). In the transfer test, participants encountered all possible symbol pairs without receiving feedback. (Bottom Right) An example participant's fixations (blue circles) on a single trial, produced by the EyeLink Data Viewer software. The orange boxes (not visible to participants) indicate the predefined areas of interest (AOIs) around the symbols and outcomes.

phase only), and specifically those that occurred within the outcome AOIs. Fixations lasting less than 50 ms (4.7%) were excluded and adjacent fixations within the same AOI were merged into a single fixation prior to analysis.

Results

Model-Free Analyses

In the learning phase, participants gradually learned to choose the reward-maximizing symbols in each choice context (Figure 2, left panel). Across participants, the average proportion of correct choices within the last 30 trials was .87 (95% CI [.83, .92]), significantly above chance (.50), $t(49) = 18.56$, $p < .001$, $d = 2.62$. In the transfer test, the average proportion of correct choices across participants was .65 (95% CI [.61, .70]), also significantly above chance (.50), $t(49) = 6.85$, $p < .001$, $d = 0.97$.

There are four categories of choice pairs in the transfer test: "Old" pairs identical to those encountered during the learning phase (AB, CD, EF, GH); "Congruent" pairs, in which the symbol with the higher EV was the higher value option in its original learning context, and the symbol with the lower EV was the lower value option in its original context (e.g., AD); "Neutral" pairs, in which both symbols were either the higher value option or the lower value option in their respective learning contexts (e.g., BD); and "Incongruent" pairs, in which the symbol with the higher EV was the lower value option in its original learning context, and the symbol with the lower EV was the higher value option in its original context (e.g., BC). As the example in the Introduction illustrated, the Incongruent pairs are critical for identifying context-dependent learning biases. For each participant, we calculated the proportion of EV-maximizing choices (accuracy) for each of the four choice categories above.

A repeated-measures ANOVA indicated that accuracy differed across the four choice categories, $F(3, 147) = 64.74$, $p < .001$, $\eta_p^2 = .57$ (Figure 2, right panel). Accuracy was higher for the Old pairs compared to the Congruent, Neutral, and Incongruent pairs averaged together, higher for the Congruent pairs compared to the Neutral and Incongruent pairs averaged together, and higher for the Neutral pairs compared to the Incongruent pairs (planned orthogonal contrasts using a Šidák correction for three tests, all ps $< .001$). Further, choice accuracy was above chance for the Old pairs (M = .86, 95% CI [.81, .90]), Congruent pairs (M = .83, 95% CI [.79, .88]), and Neutral pairs (M = .63, 95% CI [.57, .68]), but below chance for the Incongruent pairs (M = .39, 95% CI [.30, .48]). These

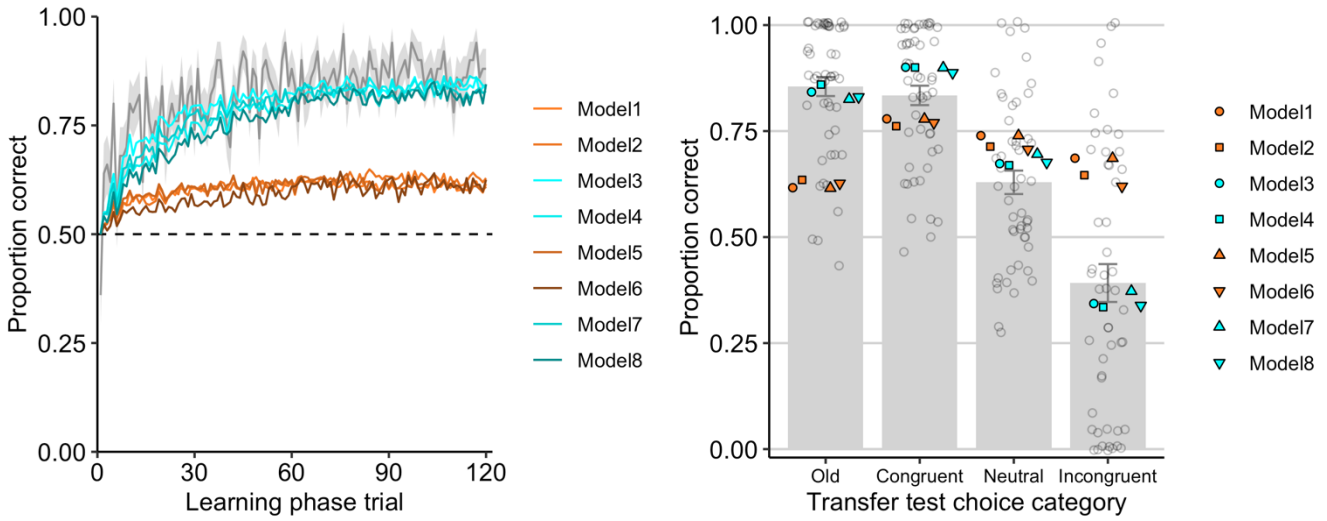

Figure 2: Proportion of correct choices across learning phase trials (left) and across the four categories of choices in the transfer test (right). In the right panel, open circles show the individual data, and bars show the mean accuracy across individuals (±1 standard error). In both panels, dark orange color is used for models that assume absolute encoding, and cyan color is used for models that assume hybrid encoding.

results support H1, demonstrating that individuals tend to prefer options that were favored in their original learning context, even when this preference contradicts EV maximization.

To test H2, we used a generalized linear mixed-effects model to predict outcome fixations (0 = no fixation, 1 = fixation) during the learning phase from trial number (divided by 120) and outcome type (0 = chosen, 1 = unchosen). The model included random intercepts and slopes to account for heterogeneity across participants. The effect of trial number was significant ($\beta = -1.33$, SE = 0.34, $p < 0.001$), with the probability of fixating on outcomes decreasing across the learning phase. Outcome type was not significant after controlling for trial number ($\beta = -0.10$, SE = 0.10, $p = 0.29$), although fixations on foregone outcomes were slightly less frequent overall than fixations on obtained outcomes. Adding an interaction between trial and outcome type did not significantly improve the model, $\chi^2(5) = 3.63$, $p = .60$. These results partially support H2, showing that the probability of fixating on outcomes decreased across the learning phase, but there was no significant difference in fixation probability between chosen and unchosen outcomes.

Model-Based Analyses

We compared several RL models fit to each participant's data. RL models maintain estimated values for the choice options that are updated in response to outcome feedback. In our models, these "Q values" were updated using one of two outcome encoding functions (absolute / hybrid), one of two learning rules (delta / decay), and assuming either full or selective attention to outcomes. Selective attention was implemented by updating an option's Q value in response to outcome feedback only if its outcome was fixated at least once on that trial.

We first describe the two different outcome encoding functions. The *absolute encoder* normalizes each outcome by the global range of outcomes experienced across all trials and contexts:

$$R_{ABS}(x_{i,t}) = \frac{x_{i,t} - \min(x_{..})}{\max(x_{..}) - \min(x_{..})}$$

where $x_{i,t}$ represents the ith outcome on trial t. The *relative encoder* normalizes the outcome relative to the minimum and maximum outcomes on the current trial, which are specific to that local context:

$$R_{REL}(x_{i,t}) = \frac{x_{i,t} - \min(x_{\cdot,t})}{\max(x_{\cdot,t}) - \min(x_{\cdot,t})}$$

In the models that assume hybrid encoding, the absolute and relative values are then combined to form an integrated subjective value for $x_{i,t}$ (see also Bavard et al., 2018; Hayes & Wedell 2023a; Molinaro & Collins, 2023):

$$R_{HYB}(x_{i,t}) = (1 - \omega_{REL}) \cdot R_{ABS}(x_{i,t}) + \omega_{REL} \cdot R_{REL}(x_{i,t})$$

The parameter ω_{REL} controls the weight given to relative outcomes. Values close to one lead to stronger context dependence, while values close to zero lead to more absolute representations that generalize across contexts (for the absolute encoding models, $\omega = 0$). For the models with selective attention, relative values are only computed on trials where the participant fixated on *both* outcomes. If only one outcome was fixated, the model computes its absolute value only.

Next, we describe the two learning rules. The delta rule tracks average reward, while the decay rule tracks cumulative reward (Don et al., 2019). More specifically, the delta rule uses the difference between experienced (R) and expected rewards (Q), weighted by the learning rate parameter α, to update the Q value for option a:

$$Q_{t+1}(a) = Q_t(a) + \omega_{fix,t} \cdot \alpha \cdot \left(R\left(x_{i,t}\right) - Q_t(a)\right)$$

The decay rule adds the value of the experienced reward to the decayed value of $Q_t(a)$, with the rate of decay controlled by the d parameter:

$$Q_{t+1}(a) = (1-d) \cdot Q_t(a) + \omega_{fix,t} \cdot R\left(x_{i,t}\right)$$

In our implementation of the decay rule, the Q values of *all* options are decayed on every trial, including options that were not seen. In the models with selective attention, an option's outcome was only integrated into the updated Q value if it received at least one fixation ($\omega_{fix,t} = 1$). Without at least one fixation ($\omega_{fix,t} = 0$), the option's Q value either remained the same (delta rule) or decayed toward zero (decay rule).[1] Finally, Q values were entered into the softmax function with a temperature parameter T to compute choice probabilities.

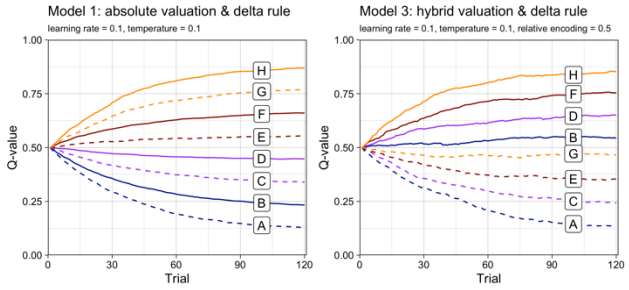

Figure 3: Mean Q values across trials for Model 1 (left) and Model 3 (right), averaged over 100 simulations with different outcome sequences. Solid (dashed) lines represent options that are locally advantaged (disadvantaged) in each context. Letters correspond to the option's expected value, from lowest (A) to highest (H).

The combination of two outcome encoding functions (absolute or hybrid), two learning rules (delta or decay), and the presence or absence of fixation-dependent updates resulted in eight candidate RL models. We simulated a subset of the models in our task to demonstrate their behavior. As shown in the left panel of Figure 3, Model 1 (absolute encoding and delta rule) predicts that learned option values align with the underlying expected values. In contrast, Model 3 (hybrid encoding and delta rule) predicts "value inversions" (Palminteri et al., 2015): Because of relative encoding, the contextually advantaged options accumulate higher Q values than the contextually disadvantaged options, despite having lower absolute EVs (Figure 3, right panel). This is what allows the model to account for suboptimal choice biases in the transfer test.

Interestingly, Model 6, which lacks a relative value mechanism and instead accumulates fixation-sampled outcomes via the decay learning rule, can also produce value inversions (Figure 4, left panel). If outcome fixations occur independently on each trial with probability p_{fix}, it can be shown that the expected Q value for option a on trial t using the decay rule is equal to:

$$E\left(Q_t(a)\right) = (1-d)^t Q_0 + p_{fix} \cdot \mu(a) \cdot \left(\frac{1-(1-d)^t}{d}\right),$$

where $\mu(a) > 0$ is the mean reward for option a. We can see that, all else equal, increasing (decreasing) the number of outcome fixations for a particular option will cause its Q value to increase (decrease) when using the decay rule.

If the outcomes from chosen options are fixated more frequently than the outcomes from unchosen options, then the contextually advantaged options B, D, and F will tend to have more outcome fixations simply by virtue of these options being chosen more often. In Model 6, this will inflate the Q values for these options relative to the contextually disadvantaged options C, E, and G, despite the latter having higher absolute rewards (Figure 4, right panel). This shows that, in theory, context-dependent choice biases can be explained without the need to posit a relative value encoding mechanism. However, as previously discussed, we found that participants fixated on obtained and forgone outcomes with similar probability (black circles in Figure 4, right panel).

The eight models were fit to each participant's data using maximum likelihood estimation. Parameters were optimized using the differential evolution algorithm from the "DEoptim" R package (Mullen et al., 2011). The Bayesian Information Criterion (BIC) was used to compare models. The results are shown in Table 1.

Based on total BIC, Model 3, which uses hybrid encoding of relative and absolute outcomes with the delta learning rule, provided the best overall fit. It was also the best fitting for 28% of participants. Overall, models incorporating hybrid encoding were the best fitting for 80% of participants, consistent with H3. Absolute encoding models did not fit the data well, as they failed to account for the high choice accuracy in the learning phase coupled with the low accuracy for Incongruent choice pairs in the transfer test (Figure 2). This was also true for the models with fixation-dependent updating. In general, adding visual fixations to the models tended to worsen the fit, although 26% of participants were best explained by a model with hybrid encoding and fixation-dependent updates.

[1] Computing $\omega_{fix,t}$ as the proportion of total outcome fixation duration on trial t provided a worse overall fit.

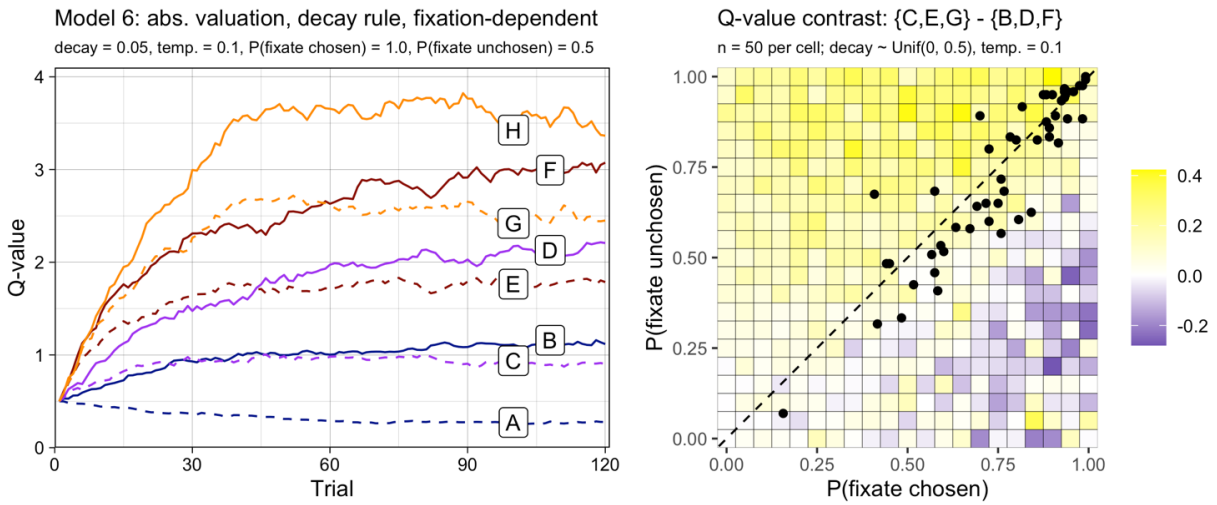

Figure 4: Value inversions (C < B, E < D, G < F) in Model 6 (left) emerge when fixations on unchosen outcomes occur less frequently than on chosen ones. The heatmap (right) shows the average difference in *Q* values between options C, E, and G and options B, D, and F across 100 simulations with different combinations of chosen and unchosen outcome fixation probabilities. Decay rates were sampled from uniform priors. Negative values indicate value inversions. Black circles show the empirical probabilities of fixating on chosen and unchosen outcomes, computed across trials, for the 50 participants.

Table 1: Model Comparisons

Model	Fixation-Dependent Updates	Valuation	Learning	Total BIC	Fit Best (%)
1	No	Absolute	Delta	13819	2 (4%)
2	No	Absolute	Decay	13501	5 (10%)
3	No	Hybrid	Delta	10713	14 (28%)
4	No	Hybrid	Decay	10800	13 (26%)
5	Yes	Absolute	Delta	13774	1 (2%)
6	Yes	Absolute	Decay	13962	2 (4%)
7	Yes	Hybrid	Delta	11069	11 (22%)
8	Yes	Hybrid	Decay	11378	2 (4%)

Discussion

We investigated the nature of context-dependent RL by comparing theory-driven models that assume different forms of outcome encoding, learning rules, and attention mechanisms. A model with hybrid (absolute and relative) outcome encoding, delta rule learning, and fixation-*independent* updating provided the best overall account of observed choice patterns.

As expected, participants consistently chose contextually favored options in the transfer test, even when this conflicted with EV maximization. This suboptimal choice bias is consistent with a growing body of research on context-dependent RL (Bavard et al., 2018; 2021; Hayes & Wedell, 2023a; Klein et al., 2017; Molinaro & Collins, 2023; Palminteri et al., 2015). Models incorporating a combination of absolute and relative outcome encoding performed the best among the eight tested, suggesting that context-dependent RL is driven by relative comparisons at the outcome encoding stage (Bavard & Palminteri, 2023; Hayes & Wedell, 2023b). We were able to rule out a plausible alternative model which relies on the accumulation of fixation-sampled outcomes (Glöckner et al., 2012) via the decay learning rule (Don et al., 2019). Importantly, this model can produce the same context-dependent choice biases without assuming any form of relative outcome encoding. However, despite being able to produce the key behavior, it was the best-fitting model for just 2 out of 50 participants.

Outcome fixations decreased across the learning phase; however, in contrast to prior studies (Ashby & Rakow, 2016; Bault et al., 2016), fixation probability did not significantly differ between chosen and unchosen outcomes. This may explain why models with fixation-dependent updating did not perform very well. There were only slightly fewer fixations on unchosen outcomes, which may not be enough of an attentional asymmetry to produce value inversions (Figure 4, right panel).

In the present study, only two outcomes were presented on every trial. However, it may be more difficult to detect relationships between fixation patterns and choice behavior using such simple displays. Future studies should explore more complex tasks with multiple alternatives and outcomes (e.g., Bavard & Palminteri, 2023; Hayes & Wedell, 2023b). In addition, modeling studies might consider integrating RL and attentional drift-diffusion models to reveal potential interactions between attention allocation and value learning (Krajbich et al., 2010).

In summary, our findings suggest that individual choices in RL tasks are driven by a hybrid process of absolute and relative outcome encoding. Although there may be a tight link between outcome fixations and choice in general (Glöckner et al., 2012), our results suggest that visual fixation patterns may not provide enough information on their own to predict whether an individual will end up learning absolute or relative values. In simplified choice environments, it seems as though the cognitive operations that transform objective outcomes into subjective values are mostly independent of overt visual attention.

References

Ashby, N. J., & Rakow, T. (2016). Eyes on the prize? Evidence of diminishing attention to experienced and foregone outcomes in repeated experiential choice. *Journal of Behavioral Decision Making, 29*(2–3), 183–193. https://doi.org/10.1002/bdm.1872

Bault, N., Wydoodt, P., & Coricelli, G. (2016). Different attentional patterns for regret and disappointment: An eye-tracking study. *Journal of Behavioral Decision Making, 29*(2–3), 194–205. https://doi.org/10.1002/bdm.1938

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature Communications*, *9*(1), 4503. https://doi.org/10.1038/s41467-018-06781-2

Bavard, S., & Palminteri, S. (2023). The functional form of value normalization in human reinforcement learning. *eLife, 12*. https://doi.org/10.7554/elife.83891

Bavard, S., Rustichini, A., & Palminteri, S. (2021). Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances, 7*(14), eabe0340.

Don, H., Otto, A., Cornwall, A., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*, 104042. 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.

Glöckner, A., Fiedler, S., Hochman, G., Ayal, S., & Hilbig, B. E. (2012). Processing differences between descriptions and experience: A comparative analysis using eye-tracking and physiological measures. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00173

Hayes, W. M., & Wedell, D. H. (2023a). Reinforcement learning in and out of context: The effects of attentional focus. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(8), 1193-1217.

Hayes, W. M., & Wedell, D. H. (2023b). Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition, 230*, 105280.

Hunter, L. E., & Daw, N. D. (2021). Context-sensitive valuation and learning. *Current Opinion in Behavioral Sciences*, *41*, 122–127. https://doi.org/10.1016/j.cobeha.2021.05.001

Klein, T. A., Ullsperger, M., & Jocham, G. (2017). Learning relative values in the striatum induces violations of normative decision making. *Nature Communications*, *8*(1), 16033. https://doi.org/10.1038/ncomms16033

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Lefebvre, G., Esmaily, J., Rezazadeh, Z., & Bahrami, B. (2024, October 9). The temporal dynamics of attentional allocation during counterfactual learning. https://doi.org/10.31234/osf.io/2y4nv

Molinaro, G., & Collins, A. G. E. (2023). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLOS Biology*, *21*(7), e3002201. https://doi.org/10.1371/journal.pbio.3002201

Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). *DEoptim: An R package for global optimization by Differential Evolution. Journal of Statistical Software, 40*(6). https://doi.org/10.18637/jss.v040.i06

Palminteri, S., & Lebreton, M. (2021). Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences, 41*, 144-151.

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Pompilio, L., & Kacelnik, A. (2010). Context-dependent utility overrides absolute memory as a determinant of choice. *Proceedings of the National Academy of Sciences*, *107*(1), 508-512.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Savage, Leonard (1954). The Foundations of Statistics. *Wiley Publications in Statistics*.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39(10), 1179–1189. http://www.jstor.org/stable/2632953

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior. Princeton University Press.*Princeton.

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.