

Attention modulates value normalization in human reinforcement learning by shaping reward encoding

Received: 28 May 2025

Accepted: 12 June 2026

Cite this article as: Cecchi, R., Gluth, S., Palminteri, S. Attention modulates value normalization in human reinforcement learning by shaping reward encoding. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-74747-w>

Romane Cecchi, Sebastian Gluth & Stefano Palminteri

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Attention modulates value normalization in human reinforcement learning by shaping reward encoding

Romane Cecchi^{1*}, Sebastian Gluth², Stefano Palminteri^{1*}

(1) Laboratoire de Neurosciences Cognitives et Computationnelles, Département d'Études Cognitives, École Normale Supérieure, Université Paris Sciences et Lettres, INSERM U960, 75005 Paris, France

(2) Department of Psychology, University of Hamburg, Hamburg, Germany

Abstract

Contextual valuation is a well-documented phenomenon in reinforcement learning, typically manifesting as range normalization in outcome representation. However, recent findings have revealed systematic deviations from this model, particularly when three options with equally spaced values are presented. In this study, we hypothesize that these distortions in outcome normalization arise from attentional processes. To test this, we conduct three experiments with 105 participants in total while simultaneously tracking their gaze position with eye-tracking. Furthermore, we systematically manipulate attention using both top-down and bottom-up approaches. These manipulations significantly increase the subjective valuation of attended options, supporting a causal role of attention in shaping value representation. To account for these effects, we develop a reinforcement learning model that integrates attentional mechanisms, wherein gaze duration directly modulates the absolute value of options prior to range normalization. This attentional range model outperforms attention-free and choice-repetition alternatives, underscoring the critical influence of attention in value computation.

Introduction

Decision-making processes are profoundly influenced by the context in which options are presented. The perceived value of an option is often shaped not solely by its intrinsic attributes but also in relation to other available alternatives. For instance, winning second place in a contest might feel rewarding until one realizes there were only two participants. This context-dependent valuation is a well-documented phenomenon, observed in both value-based decision-making scenarios, where option values are explicitly known^{1–5}, and in reinforcement learning (RL), where values are acquired through trial and error^{6–15}. This influence of context on value perception may serve as an adaptive mechanism, enabling organisms to make efficient decisions under neurobiological and environmental constraints^{16–20}. However, this same mechanism can also lead to systematic biases and seemingly irrational behaviors, particularly when choices are presented outside of their original learning context. Typically, decision-makers tend to prefer the option that was most favorable within its original learning context, even when another option has a higher absolute value but was less favorable within its own context. This behavior, which has been observed in humans and various other species, suggests that decision-makers prioritize options based on their relative value within their learning contexts, rather than their absolute value, sometimes leading to sub-optimal choices in new environments^{6,8,21–28}.

One prominent computational framework that has been proposed to account for context-dependent valuation is range normalization^{4,7,10,23,29,30}. According to this model, the subjective value of an option is rescaled based on the minimum and maximum values present in a given context. While range normalization successfully predicts subjective values in simple binary choice scenarios, it encounters limitations when additional options are introduced⁸. For example, when participants are presented with three options of evenly spaced values, the pure range normalization model predicts that the subjective value of the middle option should fall exactly halfway between the lowest and highest options. However, empirical evidence contradicts this prediction, indicating that participants tend to perceive the mid-value option as being close to the lowest one. This discrepancy indicates the presence of a non-linear bias in valuation processes, suggesting that the range model does not fully capture the cognitive mechanisms underlying subjective valuation. To address this limitation, Bavard and Palminteri⁸ introduced a weighted range normalization model, incorporating a power-transformation parameter to capture non-linearities in valuation. This refinement improves the model's predictive accuracy but remains a descriptive account, without addressing the cognitive mechanisms underlying these valuation processes.

In this study, we hypothesize that the functional form of outcome normalization could stem from attentional mechanisms that bias evidence accumulation as a function of the outcomes and expected values of available options. This hypothesis is informed by a growing body of research that has highlighted the significant role of attention in both value-based decision-making and reinforcement learning^{31–46}. Prior studies in reinforcement learning have demonstrated a bidirectional interaction between selective attentional processes and learning, where high-value

features attract attention and, in turn, attention influences how value is attributed to different features^{37,44,47,48}. Here, we propose that the bidirectional interaction between attention and learning may account for the non-linearities observed in context-dependent valuation. Specifically, we suggest that the more attention paid to an option, the closer its subjective value will align with its range normalized value.

To test this hypothesis, we conducted a series of eye-tracking experiments to investigate the relationship between attention and value normalization in reinforcement learning. The study comprised three experiments in which participants performed a choice task structured within an RL framework. During a learning phase with complete feedback, participants were asked to select one of three options with evenly spaced values. This was followed by a transfer test in which they chose between all possible pairs of options without additional feedback. The use of eye-tracking allowed us to measure attention allocation in real-time, offering insights into the cognitive processes underlying decision-making and feedback evaluation. Our results indicate that both top-down (Experiment 1) and bottom-up (Experiments 2 and 3) attentional manipulations proved effective in enhancing the subjective valuation of the mid-value options during learning, as subsequently revealed by their choices in the transfer test. These findings provide initial evidence that attention contributes to shaping subjective value.

To further account for these findings, we developed an RL model incorporating attentional mechanisms. In this model, the time spent looking at options during the choice phase directly influences their absolute value before range normalization. This attentional range model performed better than the weighted range model in capturing context-dependent preferences, highlighting the critical role of attention in shaping value representations in human reinforcement learning. Our findings offer insights into the cognitive mechanisms underlying context-dependent valuation in human decision-making, suggesting that attentional processes play a crucial role in shaping the representations that underlie learning. By integrating attentional processes into models of value normalization, we provide a more comprehensive account of how subjective value is constructed and how biases in attention allocation can influence learning.

Results

General experimental approach

To investigate the role of attention in shaping subjective value representation, we conducted three eye-tracking experiments with a total of 105 participants (35 per experiment), who performed a choice task structured within an RL framework. The experimental design consisted of two phases: a learning phase and a subsequent transfer phase (Fig. 1). During the learning phase, participants were incentivized to identify the most rewarding option in four different, interleaved contexts, each containing three options with evenly spaced values. The inclusion of three options per context was essential for examining the non-linear bias in subjective valuation, and for demonstrating the limitations of the range model in accounting for this behavior. In all three experiments, two key factors were orthogonally manipulated across the four contexts: (1) the range of possible outcomes and (2) the allocation of attention. The manipulation of the outcome range aimed to distinguish between unbiased (absolute) and context-dependent (normalized) learning processes. Meanwhile, the attention manipulation was conducted to assess the influence of attentional focus on valuation, with specific attention manipulations varying across the three experiments. Following the learning phase, participants completed a transfer test in which they made choices among all possible binary combinations of the options without receiving any additional feedback. This phase allowed us to assess how the reinforcement values acquired during the learning phase guided choices in new contexts, thereby shedding light on the interaction between attention and subjective valuation.

Experiment 1: Modulating top-down attention

The first experiment replicated the design of Bavard and Palminteri⁸, with the addition of eye-tracking to monitor participants' gaze patterns (Fig. 1a). The attention manipulation consisted in contrasting two contexts: a 100% ternary context, which served as a control with no attention manipulation, and a 50% binary context, in which attention was modulated by alternating between ternary (three-option) and binary (two-option) choice trials. In the binary trials, the high-value option was visible but blurred and not available for selection, making the mid-value option the best possible choice among the available options (Fig. 1b). This manipulation was designed to make participants actively focus on the mid-value option (top-down attention), allowing us to test whether this increased attention would lead to a more accurate subjective valuation of this option during the transfer test, compared to its counterpart in the control context.

Experiment 1: Learning phase performance

To ensure that participants were actively engaged with the task, we assessed whether the correct choice rate – defined as the probability of selecting the option with the highest expected value – exceeded chance levels during the learning phase. In the 100% ternary context, where all three options were consistently available, the chance level was 1/3. In the 50% binary context, the chance level was 1/3 in ternary trials and 1/2 in binary trials. The results indicated that the correct choice

rate was significantly above chance in all contexts (least significant (W50): $M = 0.71$, $SD = 0.27$, $t_{34} = 8.39$, $p < .001$, 95% CI = [0.62, 0.81], $d = 1.39$; two-sided, one-sample Student's t -tests; Fig. 2a). These findings confirm that participants correctly understood the task, consistently choosing the high-value option when available and the mid-value option in binary trials of the 50% binary context. Further analysis of the influence of experimental factors revealed a significant main effect of option availability, with participants showing a higher correct choice rate in the 100% ternary contexts compared to the 50% binary contexts (Δ_{100-50} : $M = 0.087$, $SD = 0.13$; $F_{1,34} = 15.4$, $p < .001$, $\eta_p^2 = .31$; two-way repeated measures ANOVA). However, no significant effect of range amplitude (i.e., the difference between the highest and the smallest outcome) was observed (wide versus narrow: $F_{1,34} = 3.93$, $p = .056$, $\eta_p^2 = .10$), nor was there a significant interaction between option availability and range amplitude ($F_{1,34} = 0.76$, $p = .39$, $\eta_p^2 = .022$)⁴⁹.

In line with our main experimental manipulation, the analysis of choice rates for each individual option revealed a significant interaction between context and option value ($F_{6,204} = 178$, $p < .001$, $\eta_p^2 = .84$; two-way repeated measures ANOVA). Post-hoc tests demonstrated that, in the 50% binary context, participants selected the mid-value option significantly more often than in the 100% ternary context (Fig. 2b; wide (Δ_{50-100}): $M = 0.48$, $SE = 0.029$, $p < .001$, 95% CI = [0.40, 0.56]; narrow (Δ_{50-100}): $M = 0.44$, $SE = 0.022$, $p < .001$, 95% CI = [0.38, 0.50]). Furthermore, within the 50% binary context, participants chose the mid-value option even more frequently than the high-value option (Fig. 2b; wide ($\Delta_{mid-max}$): $M = 0.19$, $SE = 0.042$, $p < .001$, 95% CI = [0.091, 0.29]; narrow ($\Delta_{mid-max}$): $M = 0.19$, $SE = 0.030$, $p < .001$, 95% CI = [0.12, 0.26]).

Experiment 1: Transfer phase performance

Turning to the transfer phase, we first confirmed that the correct choice rate was significantly above chance ($M = 0.85$, $SD = 0.077$, $t_{34} = 27.0$, $p < .001$, 95% CI = [0.82, 0.88], $d = 4.47$; two-sided, one-sample Student's t -tests), thereby providing evidence of successful value retrieval and generalization^{6-10,50}. Next, we analyzed the choice rate per symbol, which reflects the average frequency with which a given symbol is chosen in the transfer phase (i.e., across all possible pairings) and serves as a measure of subjective preference for that option¹³ (Fig. 2c). Our initial focus was to identify signatures of value normalization (context-dependency) and non-linear biases in subjective valuation within the 100% ternary context, which did not involve any attentional manipulation and thus served as a control. Results showed that participants chose the mid-value option in the wide context significantly less often than the high-value options in the narrow context, despite both options sharing the same objective value of 50 (Δ_{W-N} : $M = -0.44$, $SD = 0.22$, $t_{34} = -11.9$, $p < .001$, 95% CI = [-0.51, -0.36], $d = -2.75$; two-sided, paired-sample Student's t -tests; see Supplementary Fig. 1 for pairwise choice preferences). This finding suggests that participants normalized values within each context, supporting context-dependent valuation. Further analysis revealed that the choice rates for the mid-value options did not significantly differ from those of the corresponding low-value options in either context (wide: $t_{34} = -0.47$, $p = .64$, 95% CI = [-0.11, 0.066], $d = -0.10$; narrow: $t_{34} = -0.28$, $p = .78$, 95% CI = [-0.11, 0.081], $d = -0.077$; two-sided, paired-sample Student's t -tests), and Bayesian paired-sample t -tests supported the conclusion that

choice rates did not differ between option types (wide: $BF_{01} = 4.97$; narrow: $BF_{01} = 5.32$). This pattern is consistent with a non-linear processing of outcomes, in line with findings from the replicated study⁸. Given that both range normalization and non-linear biases in subjective valuation were consistently observed across the narrow and wide contexts, these contexts were combined for all subsequent analyses.

Next, we examined the effect of attention manipulation in the 50% binary context, where the high-value option was unavailable in half of the trials. Our aim was to determine whether the increased focus on the mid-value option observed during the learning phase would translate to a higher subjective valuation of these options in the transfer phase. To test this hypothesis, we fitted a weighted range normalization model separately for the 100% ternary and 50% binary contexts (Fig. 2d). Like all models tested in the present study, this model builds on a standard Q-learning architecture, in which subjective option values (Q-values) are updated from trial to trial using reward prediction errors^{51,52} and converted into choice probabilities via a softmax decision rule⁵³ (see Methods, “Computational modeling” for details). Within this framework, we used the weighted range model as the baseline because prior work shows that it closely captures participants’ behavior in this task⁸, and in our data it similarly outperformed alternative accounts – such as perseveration or psychophysical compression – which failed to reproduce the key behavioral patterns (Supplementary Fig. 2-Supplementary Fig. 4).

The model computes a subjective normalized reward R_{norm} based on the position of the objective reward R within its contextual range, followed by a non-linear transformation:

$$R_{norm} = \left(\frac{R - R_{min}}{R_{max} - R_{min}} \right)^{\omega} \quad 1$$

where R_{max} and R_{min} are the endpoints of the contextually relevant distribution, forming the range ($R_{max} - R_{min}$), and ω controls the degree of non-linear transformation (with $0 \leq \omega < +\infty$). For $\omega = 1$, the model reduces to the standard range normalization model; for $\omega < 1$, the model induces a concave deformation of the normalized outcome, leading to an overvaluation of the mid-value option; and for $\omega > 1$, it induces a convex deformation of the normalized outcome, resulting in an undervaluation of the mid-value option.

We hypothesized that an effective top-down attention manipulation would result in distinct ω parameters for the 100% ternary (ω_{100}) and 50% binary (ω_{50}) contexts. Specifically, we expected ω_{100} to be greater than 1, reflecting the non-linear processing of outcomes. In contrast, we anticipated that ω_{50} would approximate 1, reflecting near-linear scaling. The results supported these predictions: ω_{100} was significantly greater than 1 ($M = 4.67$, $SD = 2.70$, $t_{34} = 8.02$, $p < .001$, 95% CI = [3.74, 5.60], $d = 1.33$; two-sided, one-sample Student’s t-tests). In contrast, ω_{50} did not significantly differ from 1 ($M = 1.16$, $SD = 1.08$, $t_{34} = 0.90$, $p = .37$, 95% CI = [0.79, 1.54], $d = 0.15$), and Bayesian evidence favored the hypothesis that ω_{50} equals 1 ($BF_{01} = 3.79$). Furthermore, these two ω parameters were significantly different from each other ($t_{34} = 6.73$, $p < .001$, 95% CI

= [2.45, 4.56], $d = 1.66$; two-sided, paired-sample Student's t -tests), indicating that mid-value options were valued more highly in the 50% binary context compared to the full ternary context.

This result suggests that mid-value options are valued less accurately when they are presumably given less attention during learning due to the modulation of the high-value option's availability. To investigate this effect further, we next analyzed gaze data to determine whether these differences in valuation could be attributed to variations in attentional patterns.

Experiment 1: Gaze patterns

Our analysis examined how attention allocation unfolded during the learning phase and whether these patterns could account for non-linear biases in value normalization. Specifically, we analyzed fixation ratio, computed as the proportion of time spent looking at a given option relative to either (1) the total choice time (from option presentation to selection; Fig. 3a) or (2) the outcome display time (fixed at 2 seconds; Fig. 3b).

A three-way repeated measures ANOVA revealed a significant interaction between context (100% ternary vs. 50% binary), trial type (ternary vs. binary trials), and option value (min, mid, or max) across both timeframes (choice phase: $F_{2,68} = 56.0$, $p < .001$, $\eta_p^2 = .62$; outcome phase: $F_{2,68} = 27.1$, $p < .001$, $\eta_p^2 = .44$). Post-hoc analyses indicated that, in both the 100% ternary context and ternary trials of the 50% binary context, the high-value option received the most attention (Supplementary Table 1). However, in binary trials of the 50% binary context, the mid-value option received significantly more attention than the other options (Supplementary Table 1). Importantly, this shift in attentional focus was reflected in a higher fixation ratio for the mid-value option in binary trials compared to both the 100% ternary context (choice phase ($\Delta_{50B-100T}$): $M = 0.26$, $SE = 0.013$, $p < .001$, 95% CI = [0.23, 0.29]; outcome phase ($\Delta_{50B-100T}$): $M = 0.13$, $SE = 0.017$, $p < .001$, 95% CI = [0.089, 0.17]), and the ternary trials of the same context (choice phase ($\Delta_{50B-50T}$): $M = 0.20$, $SE = 0.013$, $p < .001$, 95% CI = [0.17, 0.24]; outcome phase ($\Delta_{50B-50T}$): $M = 0.13$, $SE = 0.017$, $p < .001$, 95% CI = [0.088, 0.17]). Consequently, the mid-value option received more attention overall during learning in the 50% binary context than in the 100% ternary context (Table 1), aligning with its increased valuation during the transfer phase in the same context.

In summary, these results suggest that option availability biased attention allocation, directing fixations toward the mid-value option in the 50% binary context, and that these attentional shifts may underpin non-linear value processing.

Integrating visual fixations to the range model

To formally test the hypothesis that attention modulates learning processes, we designed a modified version of the range normalization model that incorporates visual fixation data. Specifically, this approach allowed the proportion of time spent looking at an option to influence its absolute value (Equation 2) before applying the range normalization procedure (Equation 3). The model ensures that equal attention allocation across options maintains their proper positioning

within the contextual range. However, an unequal distribution of attention – where some options receive disproportionately higher or lower fixation times – can skew the relative positioning of these options within their contextual range, thereby affecting their perceived value.

To quantify the contribution of the different sources of visual information to value computation, we calculated an attention-biased option value, R_{att} , as a weighted combination of its stimulus (fix_{stim}) and outcome (fix_{out}) fixation ratios as follows:

$$R_{\text{att}} = R \cdot (\varphi \cdot \text{fix}_{\text{stim}} + (1 - \varphi) \text{fix}_{\text{out}}) \quad 2$$

Here, R denotes the objective reward value, and φ is a free parameter ($0 \leq \varphi \leq 1$) that determines the relative contribution of stimulus versus outcome fixations on value modulation. Critically, the fixation ratios fix_{stim} and fix_{out} represent the proportion of time spent looking at a given option relative to the total duration of the choice or outcome phase, respectively (with $0 \leq \text{fix}_{\text{stim/out}} \leq 1$).

Following this weighting, the attention-biased option value R_{att} was range-normalized as:

$$R_{\text{norm}} = \frac{R_{\text{att}} - \min(R_{\text{att}})}{\max(R_{\text{att}}) - \min(R_{\text{att}})} \quad 3$$

To assess how attention influenced the learning process, we tested three versions of this attentional range model. In the first version, the weight parameter φ was free to vary ($0 \leq \varphi \leq 1$), allowing a flexible combination of stimulus and outcome fixations. In the second version, φ was fixed at 1, meaning only stimulus fixations contributed to the biased option value. In the third version, φ was set to 0, so that only outcome fixations were considered. These attentional range models were compared alongside the weighted range model (Equation 1), and a standard model encoding objective reward values (“absolute” model; see Table 2 for an overview of model formulations). Model fitting was conducted jointly across the learning and transfer phases by maximizing the log-likelihood of the observed choices⁵⁴.

Experiment 1: Parameter and model comparison

The mean estimated φ parameter from the flexible attentional range model (where φ was free to vary) was significantly greater than 0.5 across participants, indicating a stronger influence of stimulus fixations (compared to outcome fixations) in explaining participants’ choices ($M = 0.65$, $SD = 0.32$, $t_{34} = 2.78$, $p = .009$, 95% CI = [0.54, 0.76], $d = 0.46$; two-sided, one-sample Student’s t -tests; Fig. 4a – see Supplementary Fig. 5a for parameter recovery analysis). This observation aligns with the model comparison results based on the Bayesian Information Criterion (BIC)⁵⁵, which identified the attentional range model relying exclusively on stimulus fixations ($\varphi = 1$) as the best fit of the data (Fig. 4b; Table 3 – see also Supplementary Fig. 5b for model recovery

analysis, Supplementary Table 2 for mean parameter estimates, and Supplementary Fig. 6 for model-based simulations).

We additionally tested alternative models in which subjective valuation was based directly on fixation ratios without range normalization (i.e., using only the fixation ratio as an estimate of subjective value). These models provided a poorer fit to the data compared to both the weighted and attentional range models (see Supplementary Table 3), highlighting the necessity of normalized outcome values for accurate behavioral predictions.

Finally, to assess whether the stage at which attention influences value computation affects model performance, we also tested a variant in which attention was applied after range normalization rather than before. Both formulations yielded highly similar results, and the “after normalization” variant kept outperforming all other models within the main model space, confirming that our overall conclusions are robust to this implementation choice (Supplementary Table 4).

Experiment 2 & 3: Modulating bottom-up attention

The results from Experiment 1 indicate that the observed transfer preferences are unlikely to arise from simple repetition of choices during learning. For example, in the 50% binary context, participants chose the mid-value option more often than the high-value option during learning (Fig. 2b), yet this did not translate into a higher preference for the mid-value option in the transfer phase. Consistent with this observation, a model implementing value changes as a function of repeated choices^{56,57} failed to capture participants’ transfer preferences and was outperformed by the weighted range normalization model (Supplementary Fig. 4). However, because attention in Experiment 1 was manipulated via changes in option availability, attentional allocation and choice patterns were inherently intertwined. Thus, although simple choice repetition cannot explain the observed changes in subjective valuation, part of the effect may still reflect choice-related dynamics rather than purely attentional mechanisms. Moreover, the similarity in fixation patterns for stimuli and outcomes (i.e., increased fixation ratio for both the mid-value stimulus and mid-value outcome; see Fig. 3) limits our ability to fully disentangle their individual contributions to subjective value computation.

To address these limitations, we conducted two follow-up experiments designed to isolate the specific influence of attention on subjective value computation. In Experiment 2, we manipulated bottom-up attention allocation by varying the saliency of the stimuli during the learning phase (Fig. 5a). In the ‘100% max’ context, the high-value stimulus was made more visually salient by increasing its brightness on each trial. In the ‘50% mid’ context, half of the trials mirrored the ‘100% max’ context (high-value trials), while in the other half, the increased brightness was applied to the mid-value stimulus instead (mid-value trials). Experiment 3 followed a similar design to Experiment 2, but the saliency manipulation was applied to outcomes rather than stimuli, irrespective of participants’ choices (Fig. 5b). These manipulations aimed to increase fixation time on the mid-value stimulus or outcome during the learning phase, without affecting option

availability. The goal was to assess whether this increased attention to the mid-value stimulus and/or outcomes would subsequently alter its subjective valuation in the transfer phase.

Experiment 2 & 3: Learning phase performance

We first examined whether saliency manipulations affected participants' ability to select the best option. In both experiments, the correct choice rate was significantly above chance level (1/3) across all contexts (least significant Exp. 2: $M = 0.89$, $SD = 0.19$, $t_{34} = 17.4$, $p < .001$, 95% CI = [0.83, 0.96], $d = 2.87$; least significant Exp. 3: $M = 0.86$, $SD = 0.19$, $t_{34} = 16.0$, $p < .001$, 95% CI = [0.79, 0.93], $d = 2.65$; two-sided, one-sample Student's t -tests; Supplementary Fig. 7). A two-way repeated measures ANOVA revealed no significant main effects or interactions of saliency or range amplitude on correct choice rates (all $p > .05$), indicating that learning performance was comparable across contexts.

We next examined choice rates for individual options to assess potential interactions between saliency contexts and option values. In contrast to Experiment 1, no significant interaction was identified between context and option value (Exp. 2: $F_{2,68} = 3.84$, $p = .057$, $\eta_p^2 = .10$; Exp. 3: $F_{2,68} = 0.70$, $p = .44$, $\eta_p^2 = .020$; two-way repeated measures ANOVA). Instead, a robust main effect of option value was observed (Exp. 2: $F_{2,68} = 1206$, $p < .001$, $\eta_p^2 = .97$; Exp. 3: $F_{2,68} = 528$, $p < .001$, $\eta_p^2 = .94$), with the high-value option being chosen more frequently than the mid- and low-value options (Supplementary Fig. 7). This indicates that choice behavior was primarily driven by option value, even when the mid-value option was made more salient.

Experiment 2 & 3: Gaze patterns

To ascertain whether the saliency manipulations had the expected effect on attention allocation, we analyzed fixation ratios during the learning phase (Fig. 5c-d). A significant three-way interaction was observed between context ('100% max' vs. '50% mid'), trial type (high-value vs. mid-value trials), and option value (min, mid, or max) across both timeframes in Experiment 2 (choice phase: $F_{2,68} = 62.3$, $p < .001$, $\eta_p^2 = .65$; outcome phase: $F_{2,68} = 86.1$, $p < .001$, $\eta_p^2 = .72$) and Experiment 3 (choice phase: $F_{2,68} = 167$, $p < .001$, $\eta_p^2 = .83$; outcome phase: $F_{2,68} = 37.4$, $p < .001$, $\eta_p^2 = .52$). In both experiments, the fixations patterns observed during the '100% max' context and the high-value trials of the '50% mid' context were similar to those observed in the 100% ternary (i.e., control) context of Experiment 1. Specifically, the high-value option received the most attention during both the choice and outcome phases (Supplementary Table 1).

However, distinct patterns emerged in mid-value trials of the '50% mid' context. During the choice phase of Experiment 2, where the attention manipulation was applied to the stimulus, participants' gaze was initially captured by the mid-value stimulus due to its increased saliency (Fig. 5c bottom). Then, over time, their attention reverted to the preferred high-value option. This initial shift in attentional focus was reflected in a significantly higher fixation ratio for the mid-value stimulus in these trials compared to both the '100% max' context ($\Delta_{50M-100H}$: $M = 0.084$, $SE = 0.018$, $p < .001$, 95% CI = [0.041, 0.13]) and the high-value trials of the same context ($\Delta_{50M-50H}$: $M = 0.085$, $SE =$

0.019, $p < .001$, 95% CI = [0.040, 0.13]). As a result, the mid-value option received more attention overall during the choice phase in the ‘50% mid’ context than in the ‘100% max’ context (Table 1). Importantly, this effect was not observed during the outcome phase, where fixation ratios for the mid-value option did not significantly differ across contexts (Table 1).

In Experiment 3, where saliency manipulation was applied to outcomes rather than stimuli, a symmetrical pattern emerged. During the outcome phase of the mid-value trials in the ‘50% mid’ context, participants’ gaze was drawn to the mid-value outcome (made more salient), but only after an initial focus on the high-value outcome (Fig. 5d bottom). This sequence aligns with the observation that participants most frequently choose the high-value option, making it the most immediate target of their post-choice gaze. Nevertheless, the increased saliency of the mid-value outcome led to a significantly higher fixation ratio for the mid-value outcome in these trials compared to both the ‘100% max’ context ($\Delta_{50M-100H}$: $M = 0.087$, $SE = 0.015$, $p < .001$, 95% CI = [0.050, 0.12]) and the high-value trials within the same context ($\Delta_{50M-50H}$: $M = 0.10$, $SE = 0.011$, $p < .001$, 95% CI = [0.072, 0.13]). Therefore, the mid-value option received significantly more attention overall during the outcome phase of the learning phase in the ‘50% mid’ context compared to the ‘100% max’ context (Table 1). Importantly, this increased attention to the mid-value outcome was not observed during the choice phase (Table 1), again contrasting with findings from Experiment 1.

Overall, these findings confirm that the saliency manipulations were effective in modulating participants’ attention in a bottom-up manner⁵⁸. In Experiment 2, the ‘50% mid’ context increased attention to the mid-value stimulus during the choice phase, while in Experiment 3, it increased attention to the mid-value outcome during the outcome phase. Importantly, these attentional shifts occurred without disrupting participants’ performance in identifying or choosing the high-value option (Supplementary Fig. 7), indicating that the saliency manipulation selectively affected attention without significantly altering choice behavior.

Experiment 2 & 3: Transfer phase performance

The next step was to assess whether attentional changes observed during the learning phase affected subsequent preferences in the transfer phase. First, we confirmed that participants’ correct choice rates remained significantly above chance in both experiments (Exp. 2: $M = 0.84$, $SD = 0.069$, $t_{34} = 29.4$, $p < .001$, 95% CI = [0.82, 0.87], $d = 4.87$; Exp. 3: $M = 0.81$, $SD = 0.061$, $t_{34} = 30.2$, $p < .001$, 95% CI = [0.79, 0.83], $d = 5.0$; two-sided, one-sample Student’s t -tests). Then, we examined the choice rates of the mid-value option across experimental contexts, as this option was the target of the attentional manipulations.

In Experiment 2, where saliency was manipulated on the stimuli, the choice rate for the mid-value option in the ‘100% max’ context did not significantly differ from that of the corresponding low-value option (Fig. 6a; $\Delta_{mid-min}$: $M = 0.046$, $SD = 0.21$, $t_{34} = 1.33$, $p = .19$, 95% CI = [-0.024, 0.12], $d = 0.35$; two-sided, paired-sample Student’s t -tests; see Supplementary Fig. 1 for pairwise choice

preferences), a pattern consistent with previously observed non-linear processing of outcomes. In the ‘50% mid’ context, however, the mid-value option was chosen significantly more often than the low-value option (Fig. 6a; $\Delta_{\text{mid-min}}$: $M = 0.11$, $SD = 0.18$, $t_{34} = 3.68$, $p < .001$, 95% CI = [0.051, 0.18], $d = 0.92$), indicating a higher subjective valuation of the mid-value option in this context. Importantly, this transfer effect was not related to simple repetition or choice-frequency effects during learning, as no reliable relationship was found between learning- and transfer-phase choice rates for the mid-value option in the 50% mid context ($\rho_{33} = -0.01$, $p = .95$; Spearman’s rank correlation). A comparison of the omega (ω) parameters from the weighted range normalization model (Equation 1) revealed a significant difference between contexts ($\Delta\omega_{100-50}$: $M = 1.52$, $SD = 3.36$, $t_{34} = 2.68$, $p = .011$, 95% CI = [0.37, 2.68], $d = 0.59$; two-sided, paired-sample Student’s t -tests), with ω_{50} being closer to 1 (Fig. 6b). This result reflects more linear processing of normalized outcomes in the ‘50% mid’ context.

In Experiment 3, where saliency was manipulated on the outcomes, the choice rate for the mid-value option did not significantly differ from that of the corresponding low-value option in the ‘100% max’ context (Fig. 6c; $\Delta_{\text{mid-min}}$: $M = 0.028$, $SD = 0.17$, $t_{34} = 0.94$, $p = .35$, 95% CI = [-0.032, 0.087], $d = 0.21$). However, although the choice rate for the mid-value option showed a slight increase in the ‘50% mid’ context, this effect was not statistically significant (Fig. 6c; $\Delta_{\text{mid-min}}$: $M = 0.058$, $SD = 0.21$, $t_{34} = 1.65$, $p = .11$, 95% CI = [-0.013, 0.13], $d = 0.41$). Similarly, fitting the weighted range normalization model revealed no significant difference in omega parameters between the two contexts (Fig. 6d; $\Delta\omega_{100-50}$: $M = 0.22$, $SD = 3.39$, $t_{34} = 0.39$, $p = .70$, 95% CI = [-0.94, 1.39], $d = 0.084$).

While the effect of the attention manipulation ($\Delta\omega_{100-50}$) was significant in Experiment 2 (stimulus saliency) but not in Experiment 3 (outcome saliency), the difference in $\Delta\omega$ between experiments was not statistically reliable ($M = 1.30$, $SD = 4.37$, $t_{68} = 1.61$, $p = .11$, 95% CI [-0.31, 2.91], $d = 0.38$; two-sided, two-sample Student’s t -test). To clarify this pattern, we performed a mixed ANOVA with Context (100% max, 50% mid) as a within-subject factor and Experiment (Exp. 2, Exp. 3) as a between-subject factor. This analysis revealed a significant main effect of Context ($F_{1,68} = 4.66$, $p = .034$, $\eta_p^2 = 0.064$) but no Context $\times$ Experiment interaction ($F_{1,68} = 2.60$, $p = .11$, $\eta_p^2 = 0.037$). Thus, across both experiments, the bottom-up attention manipulation influenced the subjective valuation of the mid-value option. Nevertheless, the pattern of results suggests a descriptive asymmetry: the effect appeared more pronounced in Experiment 2 than in Experiment 3. Although this asymmetry was not statistically confirmed by the interaction term, it is consistent with a larger impact of stimulus-directed attention on subjective valuation. To further disentangle the respective contributions of stimulus- and outcome-directed attention, we next examined inter-individual differences and conducted a fine-grained computational analysis of gaze dynamics.

Inter-individual differences in attention predict valuation distortions

While the group-level results demonstrated a clear relationship between stimulus-directed attention and subjective valuation, we investigated whether individual differences in attentional allocation could explain variations in subjective valuation and choice behavior. Specifically, we examined whether differences in fixation time on the mid-value option during learning in attention-manipulated contexts (i.e., the 50% binary or ‘50% mid’ context) predicted subsequent valuation biases in the transfer phase, as quantified by the attentional weight model parameter ω_{50} .

The results revealed a significant correlation between fixation time on the mid-value stimulus during the learning phase and the ω_{50} parameter across all three experiments (Fig. 6e; Exp. 1: $\rho_{33} = -0.47$, $p = .005$; Exp. 2: $\rho_{33} = -0.484$, $p = .0036$; Exp. 3: $\rho_{33} = -0.481$, $p = .0038$; Spearman’s rank correlation). Thus, participants who spent more time looking at the mid-value stimulus during the learning phase were more likely to choose it in the transfer phase, reflecting a more accurate subjective valuation. Conversely, participants who allocated less attention to the mid-value stimulus exhibited greater valuation distortions, undervaluing the mid-value option. Importantly, no significant correlation was observed between fixation time on outcomes and the ω_{50} parameter (Exp. 1: $\rho_{33} = 0.009$, $p = .96$; Exp. 2: $\rho_{33} = -0.074$, $p = .67$; Exp. 3: $\rho_{33} = -0.32$, $p = .060$). Together, these findings converge on a more prominent role of stimulus-directed attention in shaping subjective valuation and choice behavior.

Fine-grained computational analysis of gaze dynamics

To further investigate the computational mechanism linking attention and valuation, we compared the same five models used in Experiment 1 (absolute model, weighted range model, and the three versions of the attentional range model) in Experiments 2 and 3. In both experiments, the attentional range model with a free φ parameter (Equation 2) revealed a stronger influence of stimulus fixations compared to outcome fixations in explaining choice patterns (i.e., greater than 0.5; Exp. 2: $M = 0.83$, $SD = 0.27$, $t_{34} = 7.07$, $p < .001$, 95% CI = [0.73, 0.92], $d = 1.17$; Exp. 3: $M = 0.80$, $SD = 0.29$, $t_{34} = 6.20$, $p < .001$, 95% CI = [0.70, 0.90], $d = 1.02$; two-sided, one-sample Student’s t-tests; Fig. 4a). Additionally, the φ parameters estimated in Experiments 2 and 3 were significantly larger than those estimated in Experiment 1 (Exp. 2: $t_{68} = 2.44$, $p = .017$, 95% CI = [0.032, 0.32], $d = 0.58$; Exp. 3: $t_{68} = 2.03$, $p = .046$, 95% CI = [0.003, 0.29], $d = 0.48$; two-sided, two-sample Student’s t-tests), suggesting an even stronger dependency of value normalization on stimuli-related fixations.

Furthermore, model comparison based on BIC confirmed that the attentional range model relying exclusively on stimulus fixations ($\varphi = 1$) provided the best fit to the data (Fig. 4b; Table 3). These results further reinforce the critical role of stimulus-directed attention in reinforcement learning and subjective valuation, consistent with findings from Experiment 1.

While these model comparisons ruled out a role of outcome-related attention, the way fixation ratios were computed raises an important question. In prior analyses, fixation times have been averaged over the entire stimulus and outcome durations, implicitly assuming that all moments

within these periods contribute equally to value computation. However, it is possible that restricted time windows, such as those immediately preceding or following a decision, exert a greater influence on subjective valuation than others. For instance, given the relatively long outcome display duration (2 seconds), specific segments within this timeframe might still affect value computation.

To investigate this possibility, we refined the attentional range model by segmenting fixation data into incremental time windows. Specifically, we computed the attention-biased option value (R_{att}) by summing fixation ratios on stimuli and outcomes over specific percentages of their respective timeframes:

$$R_{att} = R \cdot (x \cdot \text{fix}_{stim} + y \cdot \text{fix}_{out}) \quad 4$$

Here, x and y represent the proportion of trial time used to calculate stimulus and outcome fixation ratios, respectively. The model was fitted incrementally with x and y varying from 0% to 100% of their respective display durations in 5% steps. For stimuli, time was considered in reverse, starting from the moment of choice and extending backward to the appearance of the stimuli. For outcomes, time was considered forward, starting from the moment of choice (i.e., the appearance of outcomes) until the disappearance of the outcomes. Note that attentional models are special cases of this ‘sliding windows’ model. For instance, the ‘stimuli’ model corresponds to [$x = 100\%$; $y = 0\%$] while the ‘outcomes’ model to [$x = 0\%$; $y = 100\%$].

The results revealed that across all three experiments, better log-likelihoods were achieved by models incorporating a larger proportion of stimulus display time compared to those emphasizing outcome time windows (Fig. 7). In particular, the best-fitting models consistently favored windows capturing most of the stimulus display period (Exp. 1: 90%; Exp. 2: 85%; Exp. 3: 80%), while incorporating only a small portion of the period immediately following the outcome display (Exp. 1: 15%; Exp. 2: 15%; Exp. 3: 10%).

To further probe the contribution of distinct temporal segments, we performed an additional analysis using non-overlapping 10% time windows across the stimulus and outcome phases (Supplementary Fig. 8). This approach tested whether time segments independent from the selection time – such as the beginning of the stimulus presentation or the end of the outcome display – could also predict subjective valuation. Results revealed that these individual 10% windows provided poorer fits than the model using the full stimulus window, suggesting that no single segment was solely driving outcome normalization. However, some windows closer to the selection time approached the performance of the full-stimulus attentional range model. Among these best-fitting windows, a larger proportion relied on stimulus-related fixations than on outcome-related ones, consistent with our main findings.

Taken together, these findings highlight the temporal dynamics of attentional influence, emphasizing that subjective value computation is best captured by stimulus fixations throughout the decision process, with minimal contribution from outcome-related attention.

ARTICLE IN PRESS

Discussion

The present study set out to investigate the role of attention in shaping subjective values during reinforcement learning (RL). Across three experiments combining attention manipulations, eye-tracking, and computational modeling, we demonstrated that attention allocation plays a critical role in explaining systematic deviations from range normalization previously observed in RL tasks. Specifically, our findings reveal that directing attention – through both top-down and bottom-up mechanisms – toward mid-value options during learning mitigates their typical undervaluation and leads to subjective valuation that better reflect their actual value and relative position within the context, as evidenced by subsequent choices. Moreover, we show that attention modulates value encoding primarily during the stimulus presentation phase, rather than during outcome evaluation, highlighting the critical role of early attentional engagement in shaping subsequent outcome processing and valuation. Together, these results provide strong evidence that attentional processes not only correlate with but causally influence value normalization in RL, offering a mechanistic explanation for previously observed non-linear distortions in subjective valuation.

Our findings replicate previous work showing that outcome values in reinforcement learning are normalized in a manner broadly consistent with the range normalization model, which linearly rescales values relative to the contextual range^{6-9,14}. At the same time, we confirmed that participants' behavior systematically deviates from the linear predictions of this model: mid-value outcomes tend to be perceived as closer to the minimum possible outcome, suggesting a systematic distortion in subjective valuation⁸. Although range normalization captures many aspects of behavior in RL, these deviations suggest that additional cognitive processes influence how subjective values are constructed during learning. To capture this pattern, our previous work introduced a weighted range model in which normalized values are further transformed by a non-linear weighting function⁸. This formulation improves descriptive accuracy and provides a coherent framework for characterizing systematic distortions in subjective valuation. However, several alternative mechanisms could, in principle, give rise to similar behavioral patterns, including psychophysical compression of outcome magnitudes^{7,59,60}, exploratory lapses during choice^{52,61}, or biases arising from past choice history (i.e., perseveration)^{56,57}. We therefore formally evaluated these empirically grounded alternatives using computational modeling (Supplementary Fig. 2-Supplementary Fig. 4). Although these mechanisms offer reasonable starting points, none reproduced the characteristic transfer-phase patterns observed in our experiments – most notably the difference between Wide and Narrow contexts for options of the same objective value, and the systematic equivalence between mid- and low-value options in transfer. Thus, these models, as currently formulated in the literature, cannot account for the observed results. By contrast, the weighted range model accommodates both the scale invariance expected from normalization and the systematic non-linear distortions observed in behavior. Importantly, while this comparison establishes the weighted range model as a descriptively adequate account within the present paradigm, it does not by itself identify the cognitive mechanisms responsible for the observed distortions. A central objective of the present study was

therefore to move beyond this descriptive account and identify the processes underlying these effects by grounding the model in a psychologically meaningful explanation.

One plausible account is that these distortions arise from how attention is allocated during learning. Extensive research in value-based decision-making has shown that attention not only reflects preferences^{62,63} but also influences value representations, with items receiving longer visual fixations being more likely to be chosen^{33,34,42,43,45,64,65}. Similarly, in risky choice, attention has been shown to systematically influence probability weighting, leading to different risk behaviors^{46,66,67}. Together, these findings suggest that attention plays a fundamental role in shaping subjective values across different decision contexts. Surprisingly, however, the relationship between visual attention and subjective value has remained relatively unexplored in reinforcement learning. Prior RL studies incorporating attentional measures primarily focused on tasks involving multi-attribute options³⁷, environments with uncertainty⁴⁴, or attentional biases toward threat⁴⁸. While these studies provided valuable insights into how attention correlates with learning, it remains unknown whether attention causally shapes subjective value representations in RL.

Building on the idea of reciprocal influences between attention and valuation observed in other domains, we hypothesized that similar mechanisms contribute to the non-linear distortions observed in RL. To test this hypothesis, we conducted a series of experiments combining eye-tracking as a measure of overt attention with attentional manipulations and computational modeling, while participants performed RL tasks inspired by prior studies. Specifically, we examined whether directing attention toward mid-value options (among three options with evenly spaced values) during learning would influence how participants valued these options later. Across experiments, we found that attention allocation during learning systematically biased subjective valuation, as reflected in participants' choices during the transfer phase.

Beyond the behavioral effects, our computational modeling results further confirmed that attention modulates value encoding during learning, thus shaping subsequent decision-making. Specifically, the model incorporating fixation data consistently outperformed both the standard range normalization model and the weighted normalization model across all three experiments, providing strong empirical support for the robustness of this effect. In this attention-modulated model, the proportion of fixations directed at an option determined how strongly newly observed outcomes updated its stored value before range normalization. This formulation is consistent with neurobiological and computational accounts proposing that attention acts as a gain-control process, amplifying the neural representation of attended stimuli before normalization occurs^{16,68}. However, when we tested the reverse implementation – where attention was applied after normalization – the results were virtually unchanged, indicating that our conclusions are robust to the precise stage at which attention is introduced in the value computation process. Overall, these findings align with previous computational models of value-based decision-making, where attention is modeled as a bias in the evidence accumulation process, leading to discounted subjective values for less-attended options^{33,34,45,64,69}.

Critically, our parameter recovery and model recovery analyses confirmed that our modeling approach was well-specified and reliably captured the underlying cognitive processes⁵⁴. Because gaze and choice are naturally correlated, one potential concern is that incorporating fixation data could introduce partial circularity – that is, if fixations merely reflected the imminent decision, this could artificially improve model fits. Importantly, several aspects of our design minimize this risk. First, fixation ratios were computed across all options and entered the model as modulators of subjective value prior to range normalization, rather than as direct predictors of choice³⁴. Second, a model relying solely on fixation data failed to outperform value-based models, confirming that gaze patterns alone are not sufficient to account for choice behavior. Third, in Experiment 2, exogenous manipulations of stimulus saliency systematically shifted fixation patterns during learning and produced corresponding changes in transfer preferences, providing causal evidence that gaze actively shapes subsequent valuation. Together, these results strengthen the conclusion that the attentional effects captured by our model are not artifacts of partial circularity, but instead reflect the genuine influence of distributed attention on subjective value construction.

Beyond confirming that attention modulates value encoding, our multi-experiment design allowed us to investigate whether this effect depends on the nature of attentional control – top-down (goal-directed) or bottom-up (stimulus-driven). In Experiment 1, we manipulated attention endogenously through the task structure: by occasionally rendering the best option unselectable within the choice set, the mid-value option became the best available alternative, naturally encouraging participants to allocate more attention to it as they selected it. In contrast, Experiments 2 and 3 relied on exogenous manipulations, enhancing the visual saliency of the mid-value option to attract attention without providing any additional task-relevant information. Crucially, these manipulations increased attention toward the mid-value option without increasing its choice frequency during learning, thereby dissociating attentional allocation from learning-phase choice repetition.

Our results revealed that both forms of attentional manipulation effectively influenced subjective valuation, providing causal evidence that attention shapes value representations during learning. However, the magnitude of the effect differed: the impact of top-down attention in Experiment 1 was stronger than that of bottom-up saliency manipulation in Experiment 2. One possible explanation is that goal-directed attention, being more tightly linked to task demands, exerts a stronger influence on value learning⁷⁰. Alternatively, the impact of stimulus-driven attention may have been reduced because participants learned to disregard the saliency manipulation once it became clear that it did not reliably predict reward – a form of habituation. This interpretation is consistent with previous findings showing that salient but uninformative cues can be ignored when they fail to signal relevant outcomes^{71–74}. Interestingly, these findings echo results from value-based decision-making tasks, where exogenous attentional manipulations – such as saliency changes or altered presentation durations – have been shown to bias choices^{38,65,75–79}, supporting a causal role of attention in shaping preferences.

Having established that attention influences value learning, we next examined when this modulation occurs – during stimulus presentation or outcome evaluation. In reinforcement learning, both phases could theoretically shape value updating: attention to stimuli might bias how information is sampled and retrieved, while attention to outcomes might influence how reward information is integrated. To test this, we compared model variants that allowed fixation data to influence value updating either during stimulus observation, outcome evaluation, or both. Specifically, we used the φ parameter to estimate the relative contribution of stimulus and outcome fixations to learning.

Across experiments, three converging results consistently pointed to attention during stimulus presentation as having a greater impact on value learning than attention during outcome observation. First, model comparison showed that the best-fitting model was the one where only stimulus fixations modulated value updating ($\varphi = 1$), outperforming both the outcome-only and the flexible models. Second, estimates of the φ parameter in the flexible model (where φ was free to vary) systematically assigned more weight to stimulus fixations, confirming that attention during option evaluation primarily drives learning. Finally, fine-grained temporal analyses of gaze data further supported this conclusion, revealing that the critical window for attentional modulation of normalized value occurs primarily during stimulus observation, with a brief refractory period immediately after outcome delivery. Taken together, these findings indicate that attention deployed during stimulus presentation sets the stage for how subsequent outcomes are processed and stored. In other words, the amount of attention allocated to an option before observing its outcome determines the strength of the subsequent value update.

Interestingly, these findings challenge the intuitive assumption that attention during outcome observation should predominantly influence the learning process, given that reward information is delivered at that stage. Instead, they suggest that it is the binding of stimuli and outcomes, established by attention at the time of choice, that shapes how values are encoded and later retrieved. This insight helps reconcile previous mixed findings in RL studies, which differed in the emphasis they placed on attention during stimulus presentation versus outcome evaluation when accounting for choice behavior^{37,44}. It further highlights that value learning depends not only on outcome processing but also critically on how attentional resources are allocated during option evaluation.

Future research could aim to develop a fully bidirectional computational framework by incorporating a generative model of attention. Rather than treating fixations as exogenous modulators of learning and choice, as done here and in previous RL studies, a more comprehensive approach would model attention as an active process that is itself learned and optimized over time. Such a framework would allow predicting not only how attention shapes value updating but also how agents learn to allocate their limited attentional resources both within and across trials. Existing models in value-based decision-making already propose reciprocal interactions between attention and choice, where attention dynamically guides and is guided by the valuation process^{31,80–83}. Similarly, some studies have developed learning models of eye movements or

visual attention^{71,84}, but these models typically do not account for how attention allocation directly shapes value learning. Extending such generative attention models into reinforcement learning contexts could clarify how learners adaptively allocate attention based on evolving value estimates, uncertainty, or task demands, offering a richer mechanistic understanding of attention's role in learning.

Taken together, our findings provide strong evidence that attention plays a causal role in shaping subjective value representations during reinforcement learning. By incorporating eye-tracking data into computational models of value learning, we move beyond descriptive accounts and toward a mechanistic understanding of how attention give rise to biases in valuation. These insights have broad implications for models of adaptive decision-making and open avenues for investigating the dynamic interplay between attention and valuation in both humans and animals.

Methods

Participants and inclusion criteria

A total of 105 healthy volunteers were recruited across three experiments ($n = 35$ per experiment; mean age = 27.6 ± 0.6 years; 55 females and 50 males, self-reported; mean years of university education = 4 ± 0.2). Participants were required to meet the following inclusion criteria: (1) age between 18 and 40 years and (2) ability to complete the experiment without glasses or contact lenses to ensure accurate eye-tracking recordings.

The compensation method followed experimental economics standards and was transparently explained to participants prior to the experiment. Each participant received a fixed amount of 13€, plus a variable bonus between 0€ and 7€, based on performance. The variable bonus was determined by the number of points gained relative to a random performance (rounded to the nearest 50 cents). A chance-level performance resulted in a 0€ bonus, while perfect performance corresponded to the maximum bonus of 7€. The total average earnings across participants were $19.0€ \pm 0.11$.

The study was approved by the INSERM Ethical Review Committee (IRB00003888). On the day of the experiment, participants were welcomed to the department's experimental platform, where the procedure was explained both orally and in writing. Written informed consent was obtained from all participants using forms sent to them in advance.

Behavioral Tasks

Participants performed an instrumental learning task adapted from previous research⁸. The experiment was conducted on a PC using a custom-built local website, which presented visual stimuli and recorded behavioral data. Responses were made using a computer mouse with the participant's preferred hand. Task instructions concluded with a three-item quiz to ensure participants understood the task. Each participant completed a training session before proceeding to the main experiment.

The full task consisted of three phases: a learning phase followed by two elicitation phases (transfer and explicit rating). Participants were unaware of the elicitation phases until they appeared during the task. The order of the elicitation phases was counterbalanced across participants. At the end of each phase, participants were shown their total points earned and the corresponding monetary equivalent.

Training

The training session was a short version of the learning phase, designed to familiarize participants with the task. It consisted of 15 trials (five trials per context across three randomly selected contexts out of four). Simple shapes replaced real stimuli to reduce cognitive load during training.

Participants' performance was evaluated based on their accuracy in choosing the option with the highest expected value. To proceed to the main experiment, participants were required to achieve at least 60% accuracy. If they failed, the training was repeated (unbeknownst to them) up to three times. If the threshold was not achieved after four attempts, participants proceeded to the main task regardless of performance (this occurred in 7 out of 105 participants). Participants could also voluntarily repeat the training if desired. During training, the experimenter remained in the booth to ensure proper functioning of the eye-tracker and clarify any questions. Once training was completed, the experimenter left, and participants completed the remainder of the experiment independently.

Learning Phase

During the learning phase, participants identified the most rewarding option in four different contexts structured in a 2×2 factorial design. Each context consisted of three options with evenly spaced values (12 options total). Two factors were orthogonally manipulated across the four contexts: (1) the range of possible outcomes (wide vs. narrow) and (2) the allocation of attention.

Each triplet of options (representing a context) was presented 45 times in an interleaved fashion, resulting in a total of 180 trials. Within each triplet, the options were associated with deterministic outcomes drawn from normal distributions with fixed means and variance. In all experiments, the means for the low-, mid-, and high-value options were [14, 32, 50] for the narrow context and [14, 50, 86] for the wide context. The variance was set at ± 4 for all options.

Options were represented by abstract stimuli generated from a 5×5 matrix, each containing 18 filled squares. Stimuli colors were selected from the HCL color scheme, with consistent chroma ($C = 55$) and luminance ($L = 70$) across options, and equally spaced hue values ($H = 0:40:360$). Stimuli positions (left, middle, or right) were randomized across trials, ensuring each option appeared an equal number of times in each position. Stimuli appeared at an angular size of approximately $4^\circ \times 4^\circ$, separated by 9.7° . The assignment of stimuli to specific context options was randomized across participants.

Participants selected an option by clicking on it. Choices were self-paced. After selection, the stimuli disappeared, and the outcomes for all three options were immediately displayed for 2 seconds. Outcomes were shown in grayscale ($C = 0$) with luminance matched to the stimuli ($L = 70$). No fixation screen was presented between trials.

In Experiment 1, attention was manipulated by varying option availability (top-down attention manipulation). In the 100% ternary context, all three options were selectable. In the 50% binary context, the high-value option was visible but unavailable for selection in 50% of the trials (interleaved). In these trials, the high-value option was blurred to maintain consistent luminance. Regardless of availability, outcomes for all options were displayed at the end of each trial.

In Experiment 2, attention was manipulated by varying stimuli saliency (bottom-up attention manipulation). All stimuli were displayed in grayscale ($C = 0$). Non-manipulated stimuli had a luminance of 70, while manipulated stimuli had a luminance of 95, making them more visually salient. In the ‘100% max’ context, the high-value stimulus was made more salient on every trial. In the ‘50% mid’ context, luminance was increased for the high-value stimulus in half the trials and for the mid-value stimulus in the other half (interleaved).

Experiment 3 followed the same design as Experiment 2, but applied saliency manipulations to outcomes instead of stimuli. Non-manipulated outcomes had a luminance of 70, while manipulated outcomes had a luminance of 95. Manipulation occurred independently of participants’ choices, with the high-value outcome made salient in the ‘100% max’ context and alternating between the high- and mid-value outcomes in the ‘50% mid’ context.

Transfer Phase

In the transfer phase, the 12 stimuli from the learning phase were presented in all possible binary combinations (66 pairs, excluding pairs formed by the same stimulus). Each pair was displayed three times, resulting in a total of 198 trials. Participants were informed that they would be presented with pairs of options from the learning phase, but the specific combinations may not have been encountered before. No feedback about outcomes was provided during this phase to avoid altering the subjective values learned in the learning phase. On each trial, participants selected one of the two options, knowing that their responses would contribute to their final payoff.

In addition to the standard choice-based elicitation phase, participants completed an explicit rating task in which they assigned subjective values to each stimulus from the learning phase. Each stimulus was presented individually three times, leading to a total of 36 trials. Participants estimated the average value of each option by adjusting a cursor on a 0 to 100 scale. This rating task was included to match the design of Bavard & Palminteri⁸ but was not analyzed in the present study.

Eye-tracking

Gaze positions were recorded throughout the task using a tower-mounted EyeLink 1000 Plus eye tracker (SR Research Ltd., Ottawa, ON, Canada), controlled via Pylink for Python 3.8. Participants’ heads were stabilized using a chin rest positioned at a fixed distance of 700 mm from a 24-inch (60.96 cm) computer screen (dimensions: 531.36 × 298.89 mm; resolution: 1920 × 1080 pixels; pixel density: 91 ppi), where visual stimuli were presented. The EyeLink system was operated in monocular mode, recording the movement of the right eye at a sampling rate of 1000 Hz. A 13-point calibration and validation procedure, provided by SR Research software, was conducted before the training phase. Drift checks were performed before the learning and transfer phases to ensure recording accuracy. If the drift check detected substantial calibration deviations, a drift correction was applied to adjust the calibration. The screen background was a uniform gray

tone (RGB: 100, 100, 100), which was maintained across the calibration procedure and the experimental task.

Gaze Data Processing

Raw gaze data output from the EyeLink system was converted into MATLAB structures using custom scripts and the FieldTrip toolbox⁸⁵ in MATLAB R2024b (The MathWorks, Inc., Natick, Massachusetts, USA).

Three areas of interest (AOIs) were defined, corresponding to the left, middle, and right option positions on the screen. Attention to a specific option was determined based on fixations falling within its AOI. The AOIs were defined as the stimulus size plus an additional margin of 1.5° of visual angle to account for eye-tracking inaccuracy and reduce false negatives⁸⁶.

Behavioral and eye-tracking analysis

Statistical analyses were performed using MATLAB Statistical Toolbox and focused on three main variables: the correct choice rate, the choice rate per option, and the fixation ratio.

The correct choice rate was defined as the probability of selecting the option with the highest expected value, calculated separately for the learning and transfer phases. The choice rate per option represented the number of times a specific option was chosen divided by the number of times it was presented. Similarly, this measure was computed separately for the learning and transfer phases. Finally, the fixation ratio quantified the proportion of time participants spent looking at an option during learning, calculated relative to the total choice time (from stimulus presentation to selection) or the outcome display time (fixed at 2 seconds).

Descriptive statistics, including mean (M), standard deviation (SD), and 95% confidence intervals (95% CI), are reported where appropriate.

Statistical tests

Participants' overall performance was assessed by comparing their correct choice rates against chance levels using two-sided, one-sample t-tests. For trials with three available options, chance level was set at 1/3, while for trials with two options, it was 1/2. Differences in choice rates between specific options were tested using two-sided, paired-sample t-tests. To account for multiple comparisons, we applied the Benjamini-Hochberg procedure to control the false discovery rate (FDR) at $\alpha = 0.05$. For all t-tests, we reported the t-statistic (t), p-value (p), and Cohen's d as a measure of effect size.

To examine the effects of experimental factors on participants' performance, we conducted two-way repeated measures ANOVA. Correct choice rates were analyzed within the framework of the 2×2 factorial design, which included the range of outcomes (wide vs. narrow) and allocation of attention (100% vs. 50% contexts). For option-specific choice rates, a 4×3 factorial design was

used in Experiment 1, incorporating the four experimental factors and option value (low, mid, or high). In Experiments 2 and 3, where wide and narrow contexts were merged, a 2×3 factorial design was used. Fixation ratios were analyzed using a three-way repeated measures ANOVA ($2 \times 2 \times 3$) to explore the effects of context (100% vs. 50%), trial type (ternary vs. binary or high-value vs. mid-value trials), and option value (low, mid, or high). Assumptions of independence, normality, and sphericity were assessed prior to conducting the analyses. Sphericity was assessed using Mauchly's test, and corrections were applied where necessary. Specifically, the Greenhouse-Geisser correction was used for epsilon estimates below 0.75, while the Huynh-Feldt correction was applied otherwise⁸⁷. For all repeated measures ANOVA, we reported the F-statistic (F), p-value (p), and partial eta squared (η_p^2) to quantify effect sizes. Post-hoc comparisons for significant effects were conducted using Tukey-Kramer tests, implemented via MATLAB's "multcompare" function. For each post-hoc comparison, we reported the estimated difference between marginal means (M), the standard error of the difference (SE), the p-value (p), and the 95% confidence intervals (95% CI)

Computational modeling

Shared framework

Participants' behavior was analyzed using variations of simple reinforcement learning models^{51,52}. These models aimed to estimate the subjective value of each option by updating their expected values across trials. At each trial t within a context c , the expected value Q of an option i was updated using a delta rule:

$$Q_t(c, i) = Q_{t-1}(c, i) + \alpha \cdot \delta_t \quad 5$$

The learning rate $\alpha \in [0,1]$ determines the extent to which new information influences the updated value estimate $Q_t(c, i)$. The prediction error at trial t , δ_t represents the difference between the subjective outcome $s(r_i)$ and the expected value $Q_{t-1}(c, i)$:

$$\delta_t = s(r_i) - Q_{t-1}(c, i) \quad 6$$

Participants' choices were modeled using a softmax decision rule⁵³, representing the probability of choosing option i over the other options in a given context c :

$$P_t(c, i) = \frac{e^{Q_t(c, i) \cdot \beta}}{\sum_{k=1}^n e^{Q_t(c, k) \cdot \beta}} \quad 7$$

Here, n is the number of options available in a trial (e.g., $n = 2$ for binary trials and $n = 3$ for ternary trials), and $\beta > 0$ is the inverse temperature parameter, which modulates choice stochasticity. Lower values ($\beta \rightarrow 0$) result in nearly equiprobable actions, while higher values ($\beta \rightarrow +\infty$) amplify differences in selection probability based on value estimates⁵².

We compared several computational models, differing only in how subjective outcomes $s(r_i)$ were computed, as described below.

Absolute model

In the absolute model, outcomes were encoded on an absolute scale, independent of the choice context. At the beginning of the learning phase ($t = 0$), initial expected values ($Q_{t=0}$) were set to 41 for all options, reflecting the empirical mean (and therefore the midpoint) of the full option set across both the wide and narrow learning contexts. Subjective values $s(r_i)$ were encoded directly as their objective reward value r_i :

$$s(r_i) = r_i, \text{ where } r_i \in [0,100] \quad 8$$

This model assumed that participants perceived outcomes in absolute terms without contextual influence.

Weighted (nonlinear) range normalization model

In the weighted range normalization model, outcomes were encoded relative to the minimum and maximum rewards available in the context using range normalization. Subjective values were further transformed by a power parameter (ω_x) to account for non-linear biases in valuation:

$$s(r_i) = \left(\frac{r_i - \min(r)}{\max(r) - \min(r)} \right)^{\omega_x} \quad 9$$

Here, $\max(r)$ and $\min(r)$ are the maximum and minimum outcomes presented in a given trial. The parameter ω_x ($0 \leq \omega < +\infty$) reflects the degree of non-linearity in valuation and was estimated separately for the two attention contexts: 100% ternary (ω_{100}) and 50% binary (ω_{50}) in Experiment 1, and ‘100% max’ (ω_{100}) and ‘50% mid’ (ω_{50}) in Experiments 2 and 3. Initial expected values were set to $Q_{t=0} = 0.5$ for all options, corresponding to the midpoint of the normalized reward range. This ensures that value estimates begin on a neutral scale, such that early learning updates are not driven by any initial preference or artificial scaling assumption.

Biases induced by attention manipulation were evaluated by comparing ω values against 1 (indicating no bias) using two-sided, one-sample t-tests. Additionally, differences between ω_{100} and ω_{50} were assessed using two-sided, paired-sample t-tests. Finally, correlations between fixation time on mid-value options during attention-manipulated contexts in the learning phase and the parameter ω_{50} were analyzed using Spearman’s rank correlation.

Attentional range normalization model

This model incorporated fixation data to adjust the encoding of subjective values. For each option i , an attention-biased option value r_{att} was calculated as:

$$r_{\text{att}} = r_i \cdot (\varphi \cdot \text{fix}_{\text{stim}} + (1 - \varphi) \text{fix}_{\text{out}}) \quad 10$$

Here, r_i represents the objective reward value, fix_{stim} and fix_{out} are the stimulus and outcome fixation ratios ($0 \leq \text{fix}_{\text{stim/out}} \leq 1$), and φ ($0 \leq \varphi \leq 1$) is a free parameter representing the relative contribution of stimulus versus outcome fixations on value modulation. The attention-biased option value r_{att} was then range-normalized to compute the subjective value $s(r_i)$:

$$s(r_i) = \frac{r_{\text{att}} - \min(r_{\text{att}})}{\max(r_{\text{att}}) - \min(r_{\text{att}})} \quad 11$$

Three versions of the attentional range model were tested: (1) φ as a free parameter, allowing a flexible combination of stimulus and outcome fixations, (2) $\varphi = 1$ (only stimulus fixations influenced subjective values), and (3) $\varphi = 0$ (only outcome fixations influenced subjective values). Q-values were updated only for selectable options on each trial, because attention to unselectable options is exogenously constrained and therefore uninformative about endogenous, value-guided learning. This constraint was only relevant for the 50% binary trials of Experiment 1, where the high-value option was blurred and unavailable for choice. Initial expected values were set to $Q_{t=0} = 0.5$ for all options.

The influence of stimulus fixations compared to outcome fixations was assessed by comparing φ values against 0.5 (indicating equal influence of both time periods) using two-sided, one-sample t-tests. Differences in φ values across the three experiments were evaluated using two-sample t-tests.

Supplementary models

INCREMENTAL WINDOW ATTENTIONAL RANGE MODEL

To explore whether all moments within the stimulus and outcome time periods equally contributed to value computation, we extended the attentional range model by segmenting fixation data into incremental time windows. The attention-biased option value r_{att} was calculated as:

$$r_{\text{att}} = r_i \cdot (x \cdot \text{fix}_{\text{stim}} + y \cdot \text{fix}_{\text{out}}) \quad 12$$

Here, x and y represent the proportion of trial time used to calculate stimulus and outcome fixation ratios, respectively. Fixation data were segmented in 5% increments, ranging from 1% to 100% of their respective display durations. For stimulus fixations, time windows were considered backward from the moment of choice to the appearance of the stimuli. For outcome fixations, time windows were considered forward from the appearance of outcomes to their disappearance.

Subjective values $s(r_i)$ were computed using the same range normalization procedure as in Equation 11. Q-values were updated only for selectable options, and initial expected values were set to $Q_{t=0} = 0.5$ for all options.

DISCRETE WINDOW ATTENTIONAL RANGE MODEL

To test whether specific moments within the stimulus or outcome display periods were more predictive of subjective valuation than others, we conducted an additional model-fitting analysis using non-overlapping 10% time windows. This analysis was designed to complement the incremental sliding window approach described above by isolating discrete time segments within each phase.

Fixation data were segmented into 10 non-overlapping windows of equal size for both the stimulus and outcome display periods: [0-10%], [10-20%], ..., [90-100%] of their respective total durations. For each model, attention-biased option values r_{att} were computed using fixation ratios from a single 10% window only, following:

$$r_{att} = r_i \cdot \text{fix}_{\text{window}} \quad 13$$

Where r_i represents the objective reward associated with option i , and $\text{fix}_{\text{window}}$ is the proportion of fixation time allocated to that option within the selected 10% window. As before, subjective values $s(r_i)$ were obtained using the range normalization procedure described in Equation 11, Q-values were updated only for selectable options, and initial expected values were set to $Q_{t=0} = 0.5$ for all options.

This procedure yielded 20 separate models per participant (10 stimulus-window models and 10 outcome-window models). To identify the best-fitting windows, we performed a repeated-measures ANOVA on log-likelihood values across all models, followed by post-hoc tests to determine which windows did not differ significantly from the model with the highest log-likelihood (among the 20 models).

FIXATION-ONLY

To evaluate whether gaze allocation already carries the information necessary for learning, we implemented fixation-only models in which subjective value was derived directly from eye-tracking data, without incorporating outcomes or range normalization (Supplementary Table 3). For each option i , an attention-derived subjective value $s(r_i)$ was calculated as:

$$s(r_i) = \varphi \cdot \text{fix}_{\text{stim}} + (1 - \varphi) \text{fix}_{\text{out}} \quad 14$$

Where, fix_{stim} and fix_{out} are the stimulus and outcome fixation ratios ($0 \leq \text{fix}_{\text{stim/out}} \leq 1$), and φ ($0 \leq \varphi \leq 1$) specifies the relative contribution of stimulus versus outcome fixations.

Three versions of the fixation-only model were tested: (1) φ as a free parameter, allowing a flexible combination of stimulus and outcome fixations, (2) a stimulus-only variant ($\varphi = 1$), and (3) an outcome-only variant ($\varphi = 0$). As in the attentional range model, Q-values were updated only for selectable options on each trial, and initial expected values were set to $Q_{t=0} = 0.5$ for all options.

ATTENTION AFTER RANGE NORMALIZATION

To assess whether the stage at which attention modulates value influences model performance, we implemented a variant in which attention modulates value following the range-normalization step (Supplementary Table 4). For each option i , the normalized outcome value was computed as:

$$r_{\text{norm}} = \frac{r_i - \min(r_i)}{\max(r_i) - \min(r_i)} \quad 15$$

Where r_i represents the objective reward value. The subjective value $s(r_i)$ was then defined as:

$$s(r_i) = r_{\text{norm}} \cdot (\varphi \cdot \text{fix}_{\text{stim}} + (1 - \varphi) \text{fix}_{\text{out}}) \quad 16$$

With fix_{stim} and fix_{out} denoting stimulus and outcome fixation ratios ($0 \leq \text{fix}_{\text{stim/out}} \leq 1$), and φ ($0 \leq \varphi \leq 1$) is a free parameter representing the relative contribution of stimulus versus outcome fixations.

Three versions of the model were tested: (1) φ as a free parameter, allowing a flexible combination of stimulus and outcome fixations, (2) a stimulus-only variant ($\varphi = 1$), and (3) an outcome-only variant ($\varphi = 0$). As in the attentional range model, Q-values were updated only for selectable options on each trial, and initial expected values were set to $Q_{t=0} = 0.5$ for all options.

POWER-LAW COMPRESSION

To evaluate whether non-linear distortions in outcome encoding could arise from psychophysical compression, we implemented a power-law compression model (Supplementary Fig. 2). In this model, subjective values were computed by applying a power transform to the objective reward r_i :

$$s(r_i) = r_i^{\gamma_x} \quad 17$$

Where $0 \leq \gamma_x \leq 1$ quantifies the degree of compression and was estimated separately for the two attention contexts: 100% ternary (γ_{100}) and 50% binary (γ_{50}) in Experiment 1, and ‘100% max’ (γ_{100}) and ‘50% mid’ (γ_{50}) in Experiments 2 and 3. Initial expected values were set to $Q_{t=0} = 41$ (mean of the full option set) for all options.

E-GREEDY

To assess whether the behavioral patterns observed in the transfer phase could be explained by exploration rather than value-based distortions, we implemented an ε -greedy model (Supplementary Fig. 3). In this model, subjective values $s(r_i)$ are encoded using standard range normalization (i.e., as in Equation 9 but without the non-linear weighting parameter). During the transfer phase, choices are assumed to be greedy with respect to learned Q-values on most trials and random otherwise. Formally, the probability of selecting option i over option j in a binary choice was defined as:

$$P_t(c, i) = \begin{cases} 1 - \varepsilon, & \text{if } Q_t(c, i) > Q_t(c, j) \\ \varepsilon, & \text{if } Q_t(c, i) < Q_t(c, j) \\ 1/2, & \text{if } Q_t(c, i) = Q_t(c, j) \end{cases} \quad 18$$

where $0 \leq \varepsilon \leq 1$ denotes the probability of making an exploratory (non-greedy) choice. Initial expected values were set to $Q_{t=0} = 0.5$ for all options.

PERSEVERATION

To evaluate whether transfer behavior could be explained by a tendency to repeat previously chosen options rather than by distortions in value representation, we implemented a perseveration model (Supplementary Fig. 4). In this model, choices depend not only on learned Q-values but also on a choice trace term $C_t(c, i)$ that captures recent selection history. The probability of selecting option i in context c at trial t (among n options; $n = 3$ in learning and $n = 2$ in transfer) was given by:

$$P_t(c, i) = \frac{e^{\beta \cdot Q_t(c, i) + \phi \cdot C_t(c, i)}}{\sum_{k=1}^n e^{\beta \cdot Q_t(c, k) + \phi \cdot C_t(c, k)}} \quad 19$$

Where β is the inverse temperature and ϕ modulates the influence of perseveration. Choice traces were updated after each trial according to:

$$\begin{aligned} C_{t+1}(c, a_t) &= C_t(c, a_t) + \tau(1 - C_t(c, a_t)), \\ C_{t+1}(c, u) &= C_t(c, u) + \tau(0 - C_t(c, u)), \forall u \neq a_t \end{aligned} \quad 20$$

with τ the accumulation/decay rate. This formulation allows the influence of past choices to build up when an option is repeatedly selected, but to decay over time when it is not chosen. Initial expected values were set to $Q_{t=0} = 41$ (mean of the full option set) for all options.

Model fitting and comparison

To identify the best-fitting computational model, parameter optimization was performed by maximizing the likelihood of observing participants' choices during both the learning and transfer

phases. Model fitting was implemented using the *fmincon* function in MATLAB software. Empirical and theoretical prior distributions were applied to constrain parameter estimates^{88,89}. Specifically, the learning rate (α), attentional weighting parameter (φ), compression parameter (γ), exploratory parameter (ε) and the choice trace learning rate (τ) were constrained by a Beta distribution (Beta(1.1,1.1)). The inverse temperature (β) and power parameter (ω) were assigned a Gamma distribution (Gamma(1.2,5)). The choice trace bias (ϕ) was given a Normal distribution ($\mathcal{N}(0,1)$).

The log model evidence was approximated by the negative half of the Bayesian Information Criterion ($-\frac{1}{2}\text{BIC}$), computed for each model. These values were submitted to a random-effects Bayesian model selection (RFX-BMS) procedure at the group level^{90,91}. RFX-BMS provides an exceedance probability (EP), which quantifies the likelihood that a given model is more frequently implemented across the population, relative to all other models in the model space.

Parameter and Model Recovery

Parameter and model recovery simulations were performed to ensure reliable parameter estimation and model distinctiveness⁵⁴. Parameter values were uniformly sampled from the range of parameters observed in participants, generating data for 1000 simulated agents. These simulations mirrored the task structure, consisting of 180 trials in the learning phase (45 per context), with deterministic outcomes drawn from normal distributions (means of [14, 32, 50] for the narrow context and [14, 50, 86] for the wide context, with a fixed variance of ± 4), and 198 trials in the transfer phase (three repetitions per pair). In the absence of a clear generative process for eye movements, fixation ratios in the attentional range models were randomly sampled between 0 and 1, ensuring that their sum across the three options was between 0 and 1. Parameter and model recovery were very good across all experiments and models (Supplementary Fig. 5).

Data availability

All raw behavioral data generated in this study are openly available on the Open Science Framework at [<https://doi.org/10.17605/OSF.IO/JWFZY>].

Code availability

All MATLAB code necessary to reproduce the analyses reported in this study is openly available on the Open Science Framework at [<https://doi.org/10.17605/OSF.IO/JWFZY>].

References

1. Busmeyer, J. R., Gluth, S., Rieskamp, J. & Turner, B. M. Cognitive and Neural Bases of Multi-Attribute, Multi-Alternative, Value-based Decisions. *Trends Cogn. Sci.* <https://doi.org/10/gf5sm8> (2019) doi:10/gf5sm8.
2. Kahneman, D. & Tversky, A. Choices, values, and frames. *Am. Psychol.* **39**, 341–350 (1984).
3. Nieuwenhuis, S. *et al.* Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage* **25**, 1302–1309 (2005).
4. Rangel, A. & Clithero, J. A. Value normalization in decision making: theory and evidence. *Curr. Opin. Neurobiol.* **22**, 970–981 (2012).
5. Tversky, A. & Simonson, I. Context-Dependent Preferences. *Manag. Sci.* **39**, 1179–1189 (1993).
6. Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G. & Palminteri, S. Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nat. Commun.* **9**, 4503 (2018).
7. Bavard, S., Rustichini, A. & Palminteri, S. Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.* **7**, eabe0340 (2021).
8. Bavard, S. & Palminteri, S. The functional form of value normalization in human reinforcement learning. *eLife* **12**, e83891 (2023).
9. Hayes, W. M. & Wedell, D. H. Effects of blocked versus interleaved training on relative value learning. *Psychon. Bull. Rev.* **30**, 1895–1907 (2023).
10. Hayes, W. M. & Wedell, D. H. Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition* **230**, 105280 (2023).
11. Hunter, L. E. & Daw, N. D. Context-sensitive valuation and learning. *Curr. Opin. Behav. Sci.* **41**, 122–127 (2021).
12. Juechems, K., Altun, T., Hira, R. & Jarvstad, A. Human value learning and representation reflect rational adaptation to task demands. *Nat. Hum. Behav.* **6**, 1268–1279 (2022).
13. Palminteri, S., Khamassi, M., Joffily, M. & Coricelli, G. Contextual modulation of value signals in reward and punishment learning. *Nat. Commun.* **6**, 8096 (2015).
14. Palminteri, S. & Lebreton, M. Context-dependent outcome encoding in human reinforcement learning. *Curr. Opin. Behav. Sci.* **41**, 144–151 (2021).
15. Spektor, M. S., Gluth, S., Fontanesi, L. & Rieskamp, J. How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychol. Rev.* **126**, 52–88 (2019).
16. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).
17. Louie, K., Glimcher, P. W. & Webb, R. Adaptive neural coding: from biological to behavioral decision-making. *Curr. Opin. Behav. Sci.* **5**, 91–99 (2015).

18. Louie, K. & Glimcher, P. W. Efficient coding and the neural representation of value. *Ann. N. Y. Acad. Sci.* **1251**, 13–32 (2012).
19. Rustichini, A., Conen, K. E., Cai, X. & Padoa-Schioppa, C. Optimal coding and neuronal adaptation in economic decisions. *Nat. Commun.* **8**, 1208 (2017).
20. Webb, R., Glimcher, P. W. & Louie, K. The Normalization of Consumer Valuations: Context-Dependent Preferences from Neurobiological Constraints. *Manag. Sci.* **67**, 93–125 (2021).
21. Freidin, E. & Kacelnik, A. Rational Choice, Context Dependence, and the Value of Information in European Starlings (*Sturnus vulgaris*). *Science* **334**, 1000–1002 (2011).
22. Klein, T. A., Ullsperger, M. & Jocham, G. Learning relative values in the striatum induces violations of normative decision making. *Nat. Commun.* **8**, 16033 (2017).
23. Padoa-Schioppa, C. Range-Adapting Representation of Economic Value in the Orbitofrontal Cortex. *J. Neurosci.* **29**, 14004–14014 (2009).
24. Pompilio, L. & Kacelnik, A. Context-dependent utility overrides absolute memory as a determinant of choice. *Proc. Natl. Acad. Sci.* **107**, 508–512 (2010).
25. Shafir, S., Waite, T. A. & Smith, B. H. Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*). *Behav. Ecol. Sociobiol.* **51**, 180–187 (2002).
26. Solvi, C. *et al.* Bumblebees retrieve only the ordinal ranking of foraging options when comparing memories obtained in distinct settings. *eLife* **11**, e78525 (2022).
27. Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
28. Vasconcelos, M., Monteiro, T. & Kacelnik, A. Context-Dependent Preferences in Starlings: Linking Ecology, Foraging and Choice. *PLOS ONE* **8**, e64934 (2013).
29. Cox, K. M. & Kable, J. W. BOLD Subjective Value Signals Exhibit Robust Range Adaptation. *J. Neurosci.* **34**, 16533–16543 (2014).
30. Soltani, A., Martino, B. D. & Camerer, C. A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence. *PLOS Comput. Biol.* **8**, e1002607 (2012).
31. Gluth, S., Kern, N., Kortmann, M. & Vitali, C. L. Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nat. Hum. Behav.* **4**, 634–645 (2020).
32. Hayes, W. M. & Wedell, D. H. Reinforcement learning in and out of context: The effects of attentional focus. *J. Exp. Psychol. Learn. Mem. Cogn.* **49**, 1193–1217 (2023).
33. Krajbich, I. Accounting for attention in sequential sampling models of decision making. *Curr. Opin. Psychol.* **29**, 6–11 (2019).
34. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* **13**, 1292–1298 (2010).
35. Krajbich, I., Lu, D., Camerer, C. & Rangel, A. The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Front. Psychol.* **3**, (2012).

36. Krajbich, I. & Rangel, A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci.* **108**, 13852–13857 (2011).
37. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron* **93**, 451–463 (2017).
38. Lim, S.-L., O'Doherty, J. P. & Rangel, A. The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. *J. Neurosci.* **31**, 13214–13223 (2011).
39. Mormann, M. M. & Russo, J. E. Does Attention Increase the Value of Choice Alternatives? *Trends Cogn. Sci.* **25**, 305–315 (2021).
40. Niv, Y. *et al.* Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
41. Sepulveda, P. *et al.* Visual attention modulates the integration of goal-relevant evidence and not value. *eLife* **9**, e60705 (2020).
42. Shimojo, S., Simion, C., Shimojo, E. & Scheier, C. Gaze bias both reflects and influences preference. *Nat. Neurosci.* **6**, 1317–1322 (2003).
43. Smith, S. M. & Krajbich, I. Gaze Amplifies Value in Decision Making. *Psychol. Sci.* **30**, 116–128 (2019).
44. Stojić, H., Orquin, J. L., Dayan, P., Dolan, R. J. & Speekenbrink, M. Uncertainty in learning, choice, and visual fixation. *Proc. Natl. Acad. Sci.* **117**, 3291–3300 (2020).
45. Ting, C.-C. & Gluth, S. Unraveling information processes of decision-making with eye-tracking data. *Front. Behav. Econ.* **3**, (2024).
46. Zilker, V. & Pachur, T. Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychol. Rev.* **129**, 949–975 (2022).
47. Niv, Y. Learning task-state representations. *Nat. Neurosci.* **22**, 1544–1553 (2019).
48. Wise, T., Michely, J., Dayan, P. & Dolan, R. J. A computational account of threat-related attentional bias. *PLOS Comput. Biol.* **15**, e1007341 (2019).
49. L'Hôtellier, M., Perez, J. & Palminteri, S. Achieving Scale-Invariant Reinforcement Learning Performance with Reward Range Normalization. Preprint at <https://doi.org/10.31234/osf.io/bjyr9> (2024).
50. Sukoupova, M., Garcia, B. & Palminteri, S. Context induces distortions in the memory representations of values: a test across multiple elicitation methods, learning modalities and architectures. Preprint at <https://doi.org/10.21203/rs.3.rs-5191803/v1> (2024).
51. Rescorla, R. & Wagner, A. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. in *Classical Conditioning II: Current Research and Theory* vol. 2 64–99 (New York: Appleton-Century-Crofts, 1972).
52. Sutton, R. S. & Barto, A. G. *Reinforcement Learning, Second Edition: An Introduction*. (MIT Press, 2018).

53. Luce, R. D. On the possible psychophysical laws. *Psychol. Rev.* **66**, 81–95 (1959).
54. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019).
55. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).
56. Wagner, B. J., Wolf, H. B. & Kiebel, S. J. Action repetition biases choice in context-dependent decision-making. *Commun. Psychol.* **3**, 177 (2025).
57. Sugawara, M. & Katahira, K. Dissociation between asymmetric value updating and perseverance in human reinforcement learning. *Sci. Rep.* **11**, 3574 (2021).
58. Posner, M. I. Orienting of Attention. *Q. J. Exp. Psychol.* **32**, 3–25 (1980).
59. Von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior*. xviii, 625 (Princeton University Press, Princeton, NJ, US, 1944).
60. Schuck-Paim, C., Pompilio, L. & Kacelnik, A. State-Dependent Decisions Cause Apparent Violations of Rationality in Animal Choice. *PLOS Biol.* **2**, e402 (2004).
61. Pisupati, S., Chartarisky-Lynn, L., Khanal, A. & Churchland, A. K. Lapses in perceptual decisions reflect exploration. *eLife* **10**, e55490 (2021).
62. Anderson, B. A. The attention habit: how reward learning shapes attentional selection. *Ann. N. Y. Acad. Sci.* **1369**, 24–39 (2016).
63. Pool, E., Brosch, T., Delplanque, S. & Sander, D. Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychol. Bull.* **142**, 79–106 (2016).
64. Cavanagh, J. F., Wiecki, T. V., Kochar, A. & Frank, M. J. Eye tracking and pupillometry are indicators of dissociable latent decision processes. *J. Exp. Psychol. Gen.* **143**, 1476–1488 (2014).
65. Bhatnagar, R. & Orquin, J. L. A meta-analysis on the effect of visual attention on choice. *J. Exp. Psychol. Gen.* **151**, 2265–2283 (2022).
66. Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O. & Hertwig, R. Prospect theory reflects selective allocation of attention. *J. Exp. Psychol. Gen.* **147**, 147–169 (2018).
67. Zilker, V. & Pachur, T. Attribute attention and option attention in risky choice. *Cognition* **236**, 105441 (2023).
68. Reynolds, J. H. & Heeger, D. J. The Normalization Model of Attention. *Neuron* **61**, 168–185 (2009).
69. Gluth, S., Spektor, M. S. & Rieskamp, J. Value-based attentional capture affects multi-alternative decision making. *eLife* **7**, e39659 (2018).
70. Vanunu, Y., Hotaling, J. M., Le Pelley, M. E. & Newell, B. R. How top-down and bottom-up attention modulate risky choice. *Proc. Natl. Acad. Sci.* **118**, e2025646118 (2021).
71. Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N. & Wills, A. J. Attention and associative learning in humans: An integrative review. *Psychol. Bull.* **142**, 1111–1140 (2016).
72. Geng, J. J., Won, B.-Y. & Carlisle, N. B. Distractor Ignoring: Strategies, Learning, and Passive Filtering. *Curr. Dir. Psychol. Sci.* **28**, 600–606 (2019).

73. Gaspelin, N., Leonard, C. J. & Luck, S. J. Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychol. Sci.* **26**, 1740–1750 (2015).
74. Gaspelin, N. & Luck, S. J. The Role of Inhibition in Avoiding Distraction by Salient Stimuli. *Trends Cogn. Sci.* **22**, 79–92 (2018).
75. Armel, K. C., Beaumel, A. & Rangel, A. Biasing simple choices by manipulating relative visual attention. *Judgm. Decis. Mak.* **3**, 396–403 (2008).
76. Colas, J. T. & Lu, J. Learning Where to Look for High Value Improves Decision Making Asymmetrically. *Front. Psychol.* **8**, (2017).
77. Gwinn, R., Leber, A. B. & Krajbich, I. The spillover effects of attentional learning on value-based choice. *Cognition* **182**, 294–306 (2019).
78. Mormann, M. M., Navalpakkam, V., Koch, C. & Rangel, A. Relative visual saliency differences induce sizable bias in consumer choice. *J. Consum. Psychol.* **22**, 67–74 (2012).
79. Towal, R. B., Mormann, M. M. & Koch, C. Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proc. Natl. Acad. Sci.* **110**, E3858–E3867 (2013).
80. Callaway, F., Rangel, A. & Griffiths, T. L. Fixation patterns in simple choice reflect optimal information sampling. *PLOS Comput. Biol.* **17**, e1008863 (2021).
81. Jang, A. I., Sharma, R. & Drugowitsch, J. Optimal policy for attention-modulated decisions explains human fixation behavior. *eLife* **10**, e63436 (2021).
82. Song, M., Wang, X., Zhang, H. & Li, J. Proactive Information Sampling in Value-Based Decision-Making: Deciding When and Where to Saccade. *Front. Hum. Neurosci.* **13**, (2019).
83. Gluth, S., Deakin, J. & Rieskamp, J. A Theory of Multi-Attribute Search and Choice. Preprint at <https://doi.org/10.31234/osf.io/3qzak> (2024).
84. Ballard, D. H., Sprague, N. & Robinson, A. Top down control accounts for gaze locations in a sidewalk navigation task. *J. Vis.* **4**, 803 (2004).
85. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
86. Orquin, J. L., Ashby, N. J. S. & Clarke, A. D. F. Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. *J. Behav. Decis. Mak.* **29**, 103–115 (2016).
87. Verma, J. P. *Repeated Measures Design for Empirical Researchers*. (John Wiley & Sons, 2015).
88. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
89. Gershman, S. J. Empirical priors for reinforcement learning models. *J. Math. Psychol.* **71**, 1–6 (2016).
90. Daunizeau, J., Adam, V. & Rigoux, L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput. Biol.* **10**, e1003441 (2014).

91. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies — Revisited. *NeuroImage* **84**, 971–985 (2014).

Funding

S.P. is supported by the European Research Council under the European Union’s Horizon 2020 research and innovation program (ERC) (RaReMem: 101043804), and the Agence Nationale de la Recherche (CogFinAgent: ANR-21-CE23-0002-02; RELATIVE: ANR-21-CE37-0008-01; RANGE: ANR-21-CE28-0024-01). The Département d’Études Cognitives is funded by the Agence Nationale pour la Recherche (ANR-17-EURE-0017, ANR-10-IDEX-0001-02). This work has received support under the Major Research Program of PSL Research University “PSL-Neuro” launched by PSL Research University and implemented by ANR (ANR-10-IDEX-0001).

Author contributions

- Conceptualization: S.P., R.C.
- Methodology: S.P., R.C.
- Investigation: R.C.
- Visualization: R.C.
- Supervision: S.P.
- Writing – original draft: R.C., S.P.
- Writing – review & editing: R.C., S.G., S.P.

Competing interests

The authors declare no competing interests.

Tables

Table 1: Post-hoc analysis of gaze patterns during learning (two-way repeated-measures ANOVA: Context × Option Value). Pairwise differences in fixation ratio between contexts (Ctx. 1 - Ctx. 2) for each option value (max, mid, min) and experiment. M = mean difference; SE = standard error of the mean; p = p-value; 95% CI = 95% confidence interval. Contexts: in Experiment 1, 100 = 100% ternary and 50 = 50% binary; in Experiments 2-3, 100 = 100% max and 50 = 50% mid.

Exp.	Opt.	Ctx. 1	Ctx. 2	Choice phase				Outcome phase			
				M	SE	p	95% CI	M	SE	p	95% CI
1	Max	100	50	0.20	0.011	< .001	[0.18 0.23]	0.074	0.010	< .001	[0.053 0.094]
	Mid	100	50	-0.16	0.010	< .001	[-0.18 -0.14]	-0.065	0.010	< .001	[-0.086 -0.044]
	Min	100	50	-0.048	7.10^{-3}	< .001	[-0.063 -0.034]	-8.10^{-3}	7.10^{-3}	.26	[-0.023 6.10^{-3}]
2	Max	100	50	0.038	0.011	.002	[0.015 0.061]	3.10^{-3}	8.10^{-3}	.71	[-0.014 0.020]
	Mid	100	50	-0.042	0.010	< .001	[-0.062 -0.022]	-5.10^{-3}	6.10^{-3}	.39	[-0.018 7.10^{-3}]
	Min	100	50	3.10^{-3}	4.10^{-3}	.50	[- 6.10^{-3} 0.012]	7.10^{-3}	7.10^{-3}	.31	[- 7.10^{-3} 0.021]
3	Max	100	50	2.10^{-3}	0.011	.86	[-0.021 0.025]	0.016	8.10^{-3}	.061	[- 8.10^{-4} 0.033]

Mid	100	50	$6 \cdot 10^{-4}$	0.010	.96	[-0.020 0.021]	-0.037	0.011	.003	[-0.060 -0.014]
Min	100	50	$-2 \cdot 10^{-3}$	$7 \cdot 10^{-3}$	.80	[-0.016 0.013]	0.019	$8 \cdot 10^{-3}$	.021	[$3 \cdot 10^{-3}$ 0.035]

Table 2 | Summary of computational models. Mathematical formulation of the attentional modulation of reward value and the resulting subjective value function $s(r_i)$ for each model. r_i : objective reward; $s(r_i)$: subjective value; r_{att} : attention-weighted option value; ω_x : degree of non-linearity in valuation (estimated separately for the two attention contexts); φ : relative contribution of stimulus vs. outcome fixations; fix_{stim} and fix_{out} : fixation ratios during the stimulus and outcome phases, respectively.

Model	Attentional modulation	Subjective value function $s(r_i)$
Absolute	None	$s(r_i) = r_i$, where $r_i \in [0, 100]$
Weighted range (ω)	None	$s(r_i) = \left(\frac{r_i - \min(r)}{\max(r) - \min(r)} \right)^{\omega_x}$
Attentional range (stimuli + outcomes)	$r_{att} = r_i \cdot (\varphi \cdot fix_{stim} + (1 - \varphi) fix_{out})$	$s(r_i) = \frac{r_{att} - \min(r_{att})}{\max(r_{att}) - \min(r_{att})}$
Attentional range (stimuli)	$r_{att} = r_i \cdot fix_{stim}$	
Attentional range (outcomes)	$r_{att} = r_i \cdot fix_{out}$	

Table 3: Model comparison across Experiments. Mean $\pm$ standard deviation (SD) of Bayesian Information Criterion (BIC) scores for each computational model tested. Lower BIC values indicate better model fit. The best model (i.e., with the lowest BIC) in each experiment is highlighted in bold.

Model	Exp. 1	Exp. 2	Exp. 3
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD
Absolute	391 $\pm$ 112	334 $\pm$ 113	408 $\pm$ 93
Weighted range (ω)	347 $\pm$ 131	264 $\pm$ 110	324 $\pm$ 125
Attentional range (stimuli + outcomes)	336 $\pm$ 125	261 $\pm$ 104	313 $\pm$ 119
Attentional range (stimuli)	336 $\pm$ 122	258 $\pm$ 103	309 $\pm$ 118
Attentional range (outcomes)	374 $\pm$ 121	310 $\pm$ 104	363 $\pm$ 115

Figure Legends

Fig. 1 | Task design. In all phases, participants were instructed to identify the most rewarding option. **(a) Learning contexts.** Illustration of the four choice contexts used during the learning phase. Each context consisted of three options with evenly spaced values and orthogonally manipulated two factors: outcome range (wide vs. narrow) and attention allocation (100% ternary vs. 50% binary). In the 100% ternary context, all trials involved three available options. In the 50% binary context, half of the trials were binary – with the high-value option blurred and unavailable for selection – resulting in an equal mix of ternary (three-option) and binary (two-option) trials. Light-grey dashed lines indicate the proportion of ternary and binary trials within each context. **(b) Learning trials.** Participants selected one option by clicking on it; choices were self-paced. Upon selection, the stimuli disappeared, and outcomes – drawn from non-overlapping normal distributions – were displayed for all three options for 2 seconds. Eye-tracking data were recorded during both the choice and outcome phases. Heatmaps illustrate normalized fixation densities for a single representative trial from one participant (for illustration only; note that stimuli were presented in color during the actual task). Option positions (left, middle, right) were randomized on each trial, and the spacing between stimuli was greater than shown here (see “Methods” for actual dimensions). **(c) Transfer trials.** Participants chose between two

previously learned options. No feedback was provided (indicated by a crossed eye icon displayed for 0.5s), allowing us to assess subjective valuation and generalization. **(d) Stimulus presentation across phases.** During learning, participants made choices within triplets of options specific to each context. In the transfer phase, all 12 stimuli from the learning phase were recombined into all possible binary pairs, enabling a context-independent assessment of preferences. **(e) Experimental timeline.** Schematic representation of trials presentation over time. During the learning phase, contexts were interleaved, with each trial corresponding to one of the four possible contexts (color-coded). In the transfer phase, all pairwise combinations of the 12 stimuli were presented, irrespective of their original learning context.

Fig. 2 | Experiment 1 behavior (top-down attention manipulation). **(a) Learning correct choice rate.** Correct choice rate during the learning phase as a function of the choice context. **(b) Cumulative learning choice count.** Cumulative number of choices per option across trials during the learning phase, illustrating the impact of the attention manipulation on choice behavior. **(c) Transfer choice rate.** Choice rate per option during the transfer phase, reflecting subjective valuation. **(d) Distortion of normalized values.** Distortions captured by the estimated omega (ω) parameters from the weighted range normalization model, fitted separately for the 100% ternary and 50% binary contexts. Each gray line represents the subjective value transformation of a single participant based on their best-fitting power parameters (ω_{100} and ω_{50}). The vertical and horizontal dashed lines illustrate how the mid-value option (normalized value = 0.5) is distorted by the non-linear transformation in each attention context, while the diagonal black dashed line represents the standard (linear) range normalization case ($\omega = 1$). In (a) and (c): Dots represent individual participant averages ($n = 35$). Violin plots show the distribution of these averages; black lines indicate the mean, and shaded areas represent the standard error of the mean (SEM; darker) and 95% confidence intervals (lighter). In (b) and (d): Colored lines represent the mean across participants; shaded areas indicate SEM. W: wide outcome range; N: narrow outcome range.

Fig. 3 | Gaze patterns during the learning phase of Experiment 1. **(a) Choice phase.** Average fixation ratios across trials (left) and time course of fixation ratios (right) for each option during the choice phase, time-locked to the onset of option presentation. **(b) Outcome phase.** Average fixation ratios across trials (left) and time course of fixation ratios (right) for each option during the outcome phase, time-locked to the onset of outcome presentation. Dots represent individual participant averages ($n = 35$). Violin plots show the distribution of these averages; black lines indicate the mean, and shaded areas represent the standard error of the mean (SEM; darker) and 95% confidence intervals (lighter). In the time courses: Colored lines represent the mean fixation ratio across participants, smoothed with a 100ms moving average for visualization purposes only; shaded areas indicate SEM.

Fig. 4 | Computational modeling results across experiments. **(a)** Estimated phi (ϕ) parameter from the attentional range model combining stimulus and outcome fixations. The ϕ parameter quantifies the relative weight of attention toward stimuli versus outcomes in modulating subjective value, with values above 0.5 indicating a stronger influence of stimulus fixations. Dots represent individual participant estimates ($n = 35$ per experiment). Violin plots show the distribution of these averages; black lines indicate the mean, and shaded areas represent the standard error of the mean (SEM) and 95% confidence intervals. **(b)** Bayesian model comparison showing estimated model frequencies and exceedance probabilities (EP) across the model space. The model relying exclusively on stimulus fixations provides the best fit, outperforming models incorporating outcome fixations as well as non-attentional models. Error bars represent the standard deviation of the posterior Dirichlet distribution over estimated model frequencies.

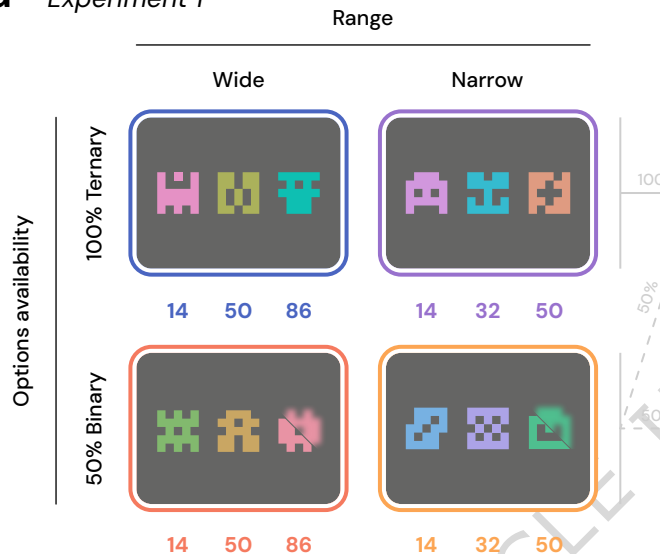
Fig. 5 | Task design and gaze data for Experiments 2 and 3 (bottom-up attention manipulations). **(a) Experiment 2 design** (stimulus saliency manipulation). During learning, the luminance of specific stimuli was increased to capture attention across two contexts. In the 100% max context, the high-value stimulus was always made more visually salient, whereas in the 50% mid context, half of the trials mirrored this manipulation (high-value trials), and in the other half, increased brightness was instead applied to the mid-value stimulus (mid-value trials). Light-grey dashed lines indicate the distribution of high-value and mid-value trials within each context. **(b) Experiment 3 design** (outcome saliency manipulation). Same design as Experiment 2, but the saliency manipulation was applied to outcome displays rather than stimuli. In both experiments, saliency was manipulated independently of participants' choices. **(c) Experiment 2 gaze data.** Average fixation ratios across trials (left) and time course of fixation ratios (right) for each option during the choice phase, time-locked to the onset of option presentation. **(d) Experiment 3 gaze data.** Average fixation ratios across trials (left) and time course of fixation ratios (right) for each option during the outcome phase,

time-locked to the onset of outcome presentation. Dots represent individual participant averages ($n = 35$ in both experiments). Violin plots show the distribution of these averages; black lines indicate the mean, and shaded areas represent the standard error of the mean (SEM; darker) and 95% confidence intervals (lighter). In the time courses: Colored lines represent the mean fixation ratio across participants, smoothed with a 100ms moving average for visualization purposes only; shaded areas indicate SEM.

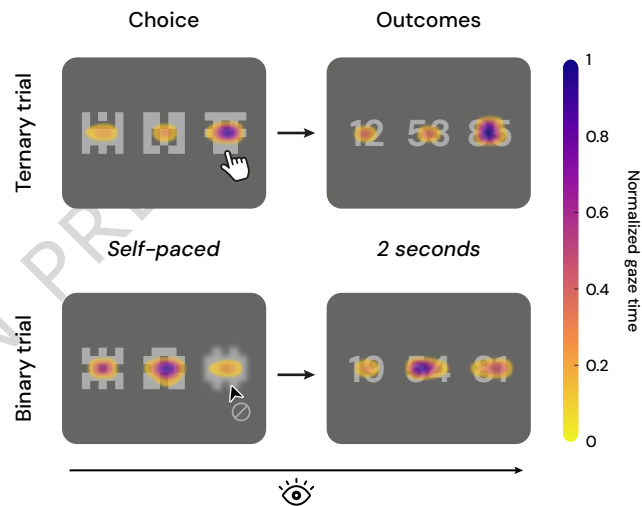
Fig. 6 | Transfer behavior in Experiments 2 and 3 and inter-individual differences. (a) *Experiment 2 transfer choice rate.* Choice rate per option during the transfer phase, illustrating increased subjective valuation of the mid-value option following stimulus saliency manipulation in the ‘50% mid’ context ($p < .001$, two-sided paired-sample Student’s t-test). (b) *Experiment 2 distortion of normalized values.* (Left) Distortions captured by the estimated omega (ω) parameters from the weighted range normalization model, fitted separately for the ‘100% max’ (purple) and ‘50% mid’ (orange) contexts. (Right) Difference between ω estimates in the two contexts ($\omega_{100} - \omega_{50}$), showing a significant increase toward linear valuation (ω closer to 1) in the ‘50% mid’ context ($p = .011$, two-sided paired-sample Student’s t-test). (c-d) Same plots as (a) and (b), respectively, for Experiment 3, where outcome saliency manipulation did not produce significant effects. (e) *Inter-individual differences.* Correlation between the proportion of fixation time allocated to the mid-value stimulus during learning (in the 50% binary or ‘50% mid’ context) and the ω_{50} parameter. Significant negative correlations indicate that greater attention to the mid-value option during learning is associated with reduced valuation distortions – that is, subjective valuations more closely reflecting the actual value (Exp. 1: $p = .005$, Exp. 2: $p = .004$, Exp. 3: $p = .004$; Spearman’s rank correlation). Dots represent individual participant averages ($n = 35$ per experiment). In (a-d), violin plots show the distribution of these averages; black lines indicate the mean, and shaded areas represent the standard error of the mean (SEM; darker) and 95% confidence intervals (lighter). In (b) and (d) (left), colored lines represent the mean across participants, with shaded areas indicating SEM. In (e), shaded areas indicate 95% confidence intervals. * $p < .05$; ** $p < .01$; *** $p < .001$; n.s.: not significant.

Fig. 7 | Temporal dynamics of attention influence on subjective valuation. (a) *Testing fixation contributions over time.* Log-likelihood differences between models using different proportions of the stimulus and outcome display windows to compute fixation ratios, relative to the attentional range model using only stimulus fixations. Higher values (pink) indicate better model fits relative to the stimulus-only model. Black squares denote the best-fitting model for each experiment. For stimulus-related fixations, negative percentages indicate that fixation windows used to compute fixation ratios are defined backward in time from the moment of choice (0) toward stimulus onset. (b) *Timing of fixation contributions.* Time courses of fixation ratios, smoothed with a 3% moving average for visualization purpose only. Time is expressed as a percentage of the total trial duration to allow comparison across trials of varying lengths. Best-fitting time windows (highlighted in pink) indicate that subjective value computation relies predominantly on fixations during the stimulus display period, with minimal contribution from outcome-related fixations. In other words, optimal model fits are achieved by incorporating most of the stimulus viewing time and only a brief period of the outcome display. Lines represent the mean across participants, with shaded areas indicating the standard error of the mean (SEM).

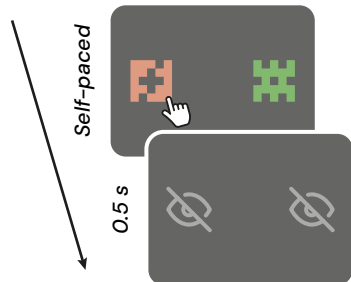
a *Experiment 1*



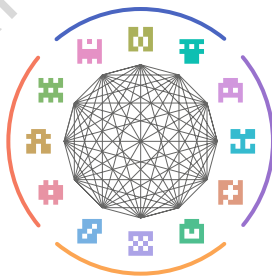
b *Learning trials*



C Transfer trials

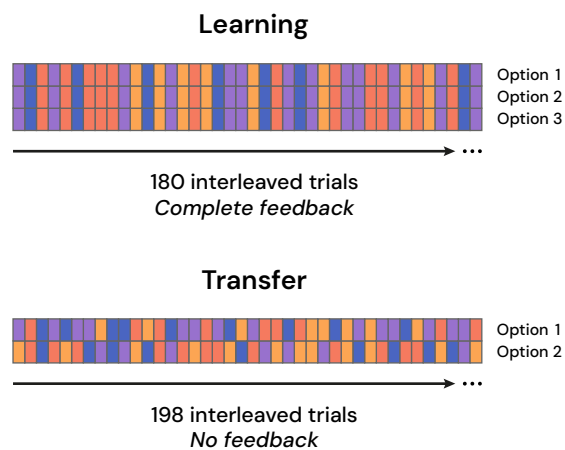


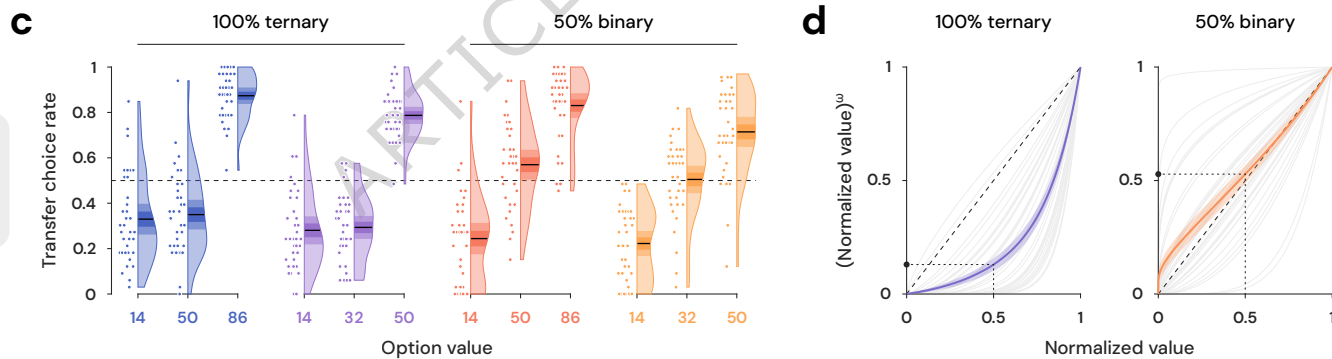
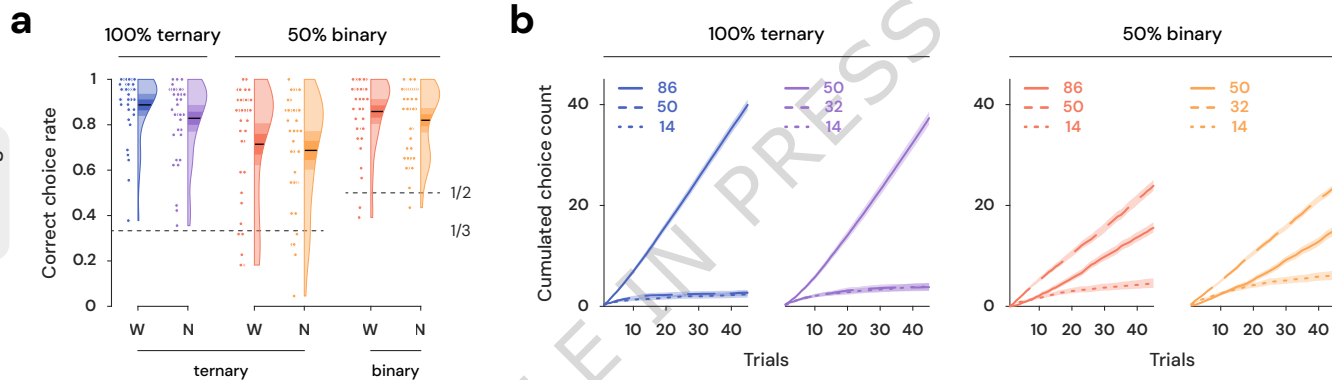
d Learning (outside)
4 contexts (45x each)



Transfer (inside)
66 pairs (3x each)

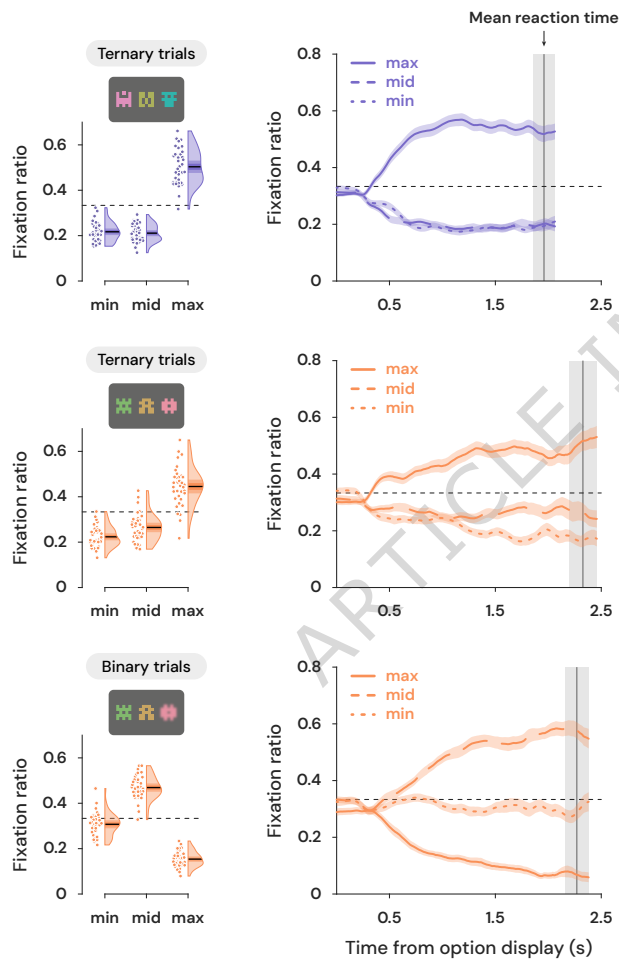
e



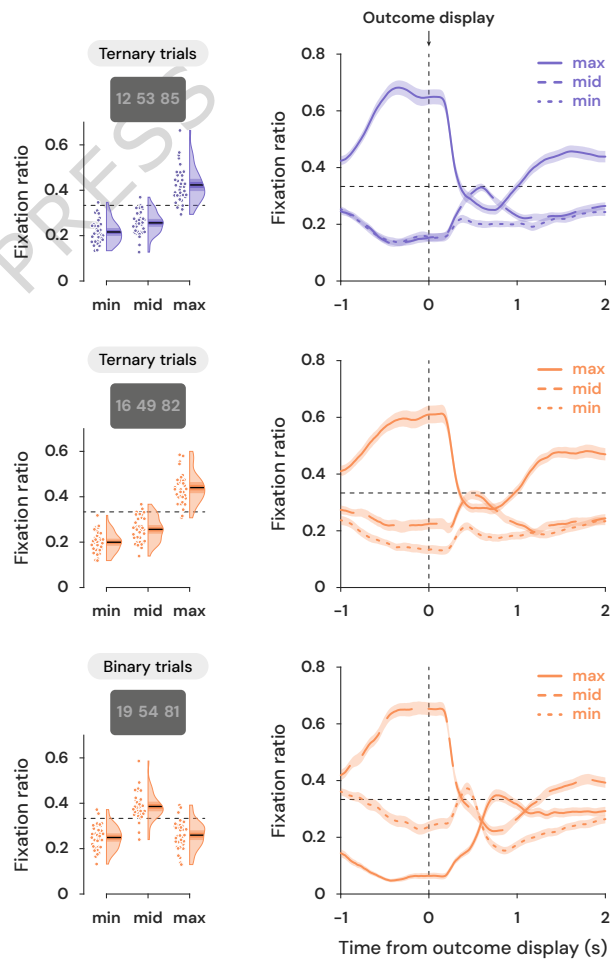


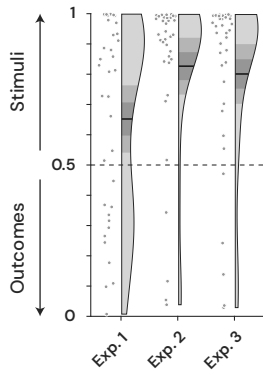
a

Choice phase

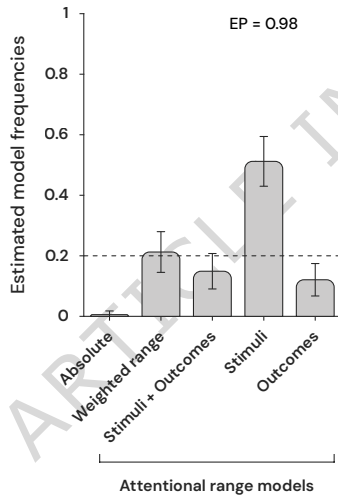
**b**

Outcome phase

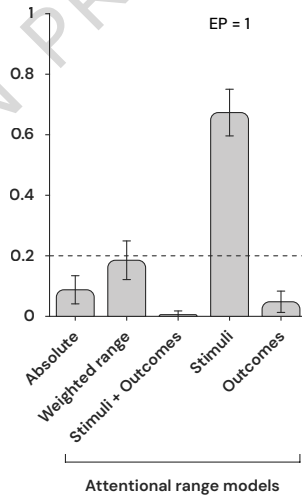


a ϕ weight**b**

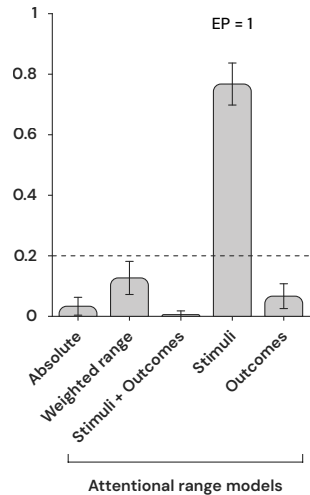
Experiment 1

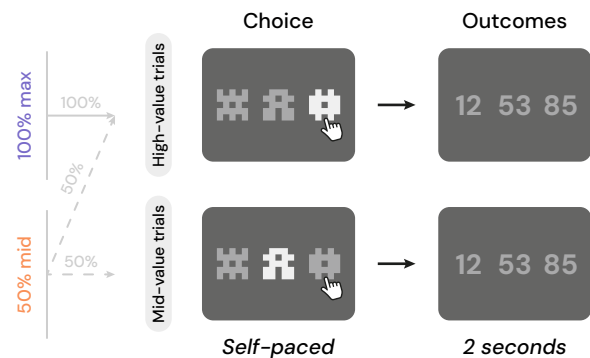
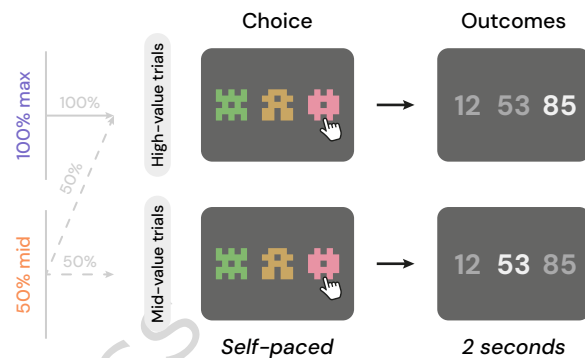
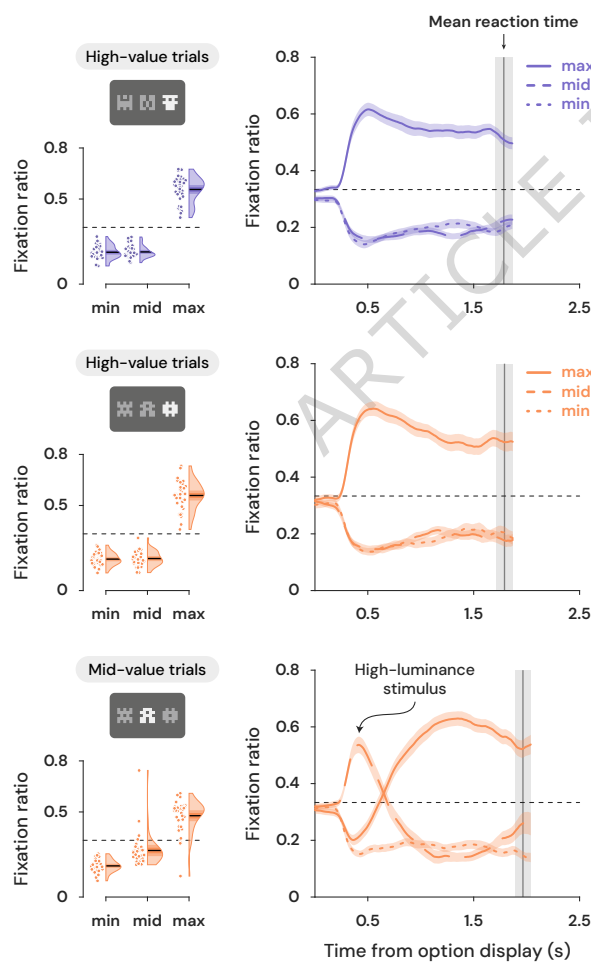
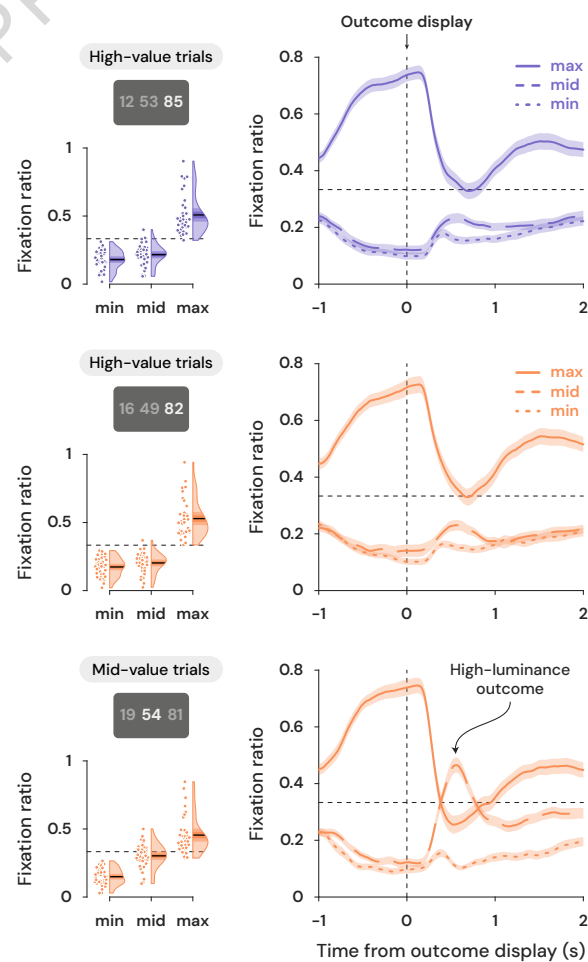


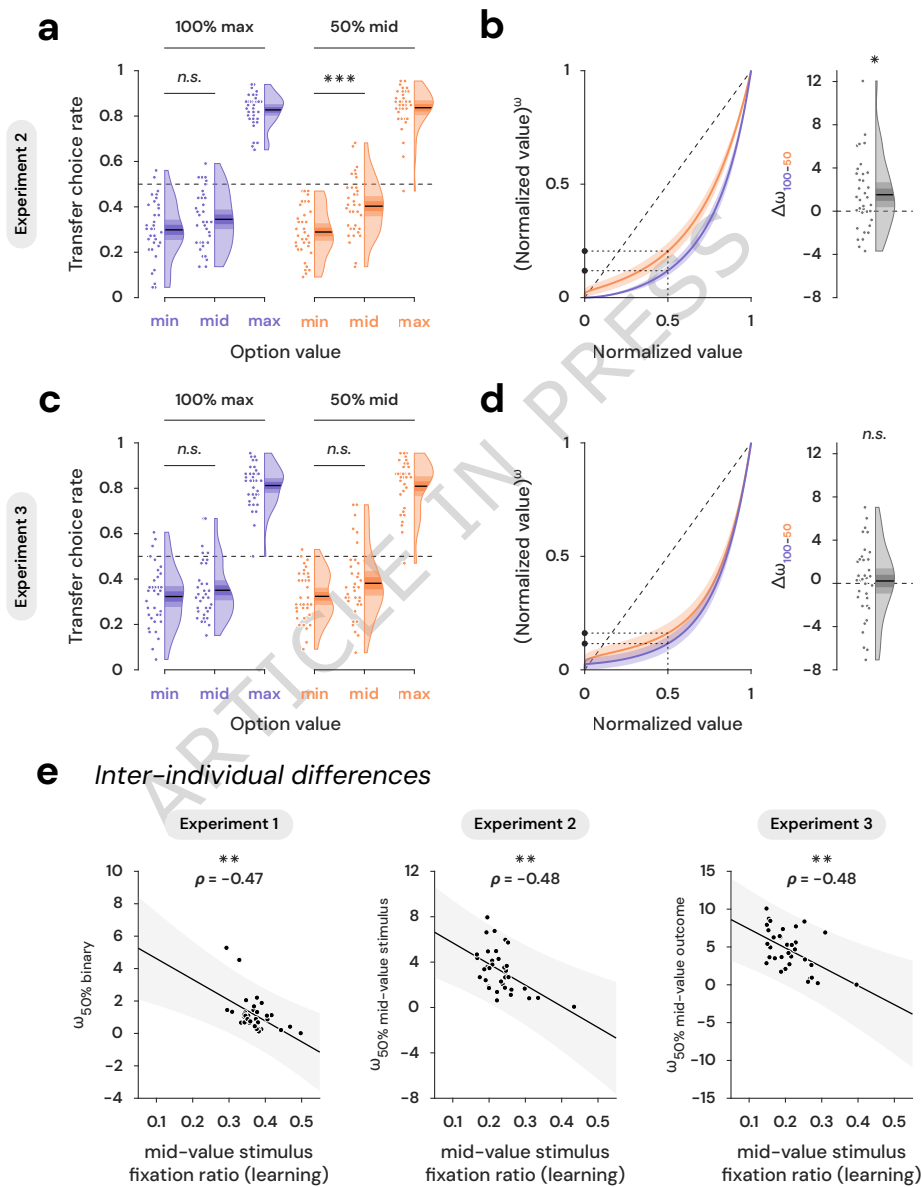
Experiment 2

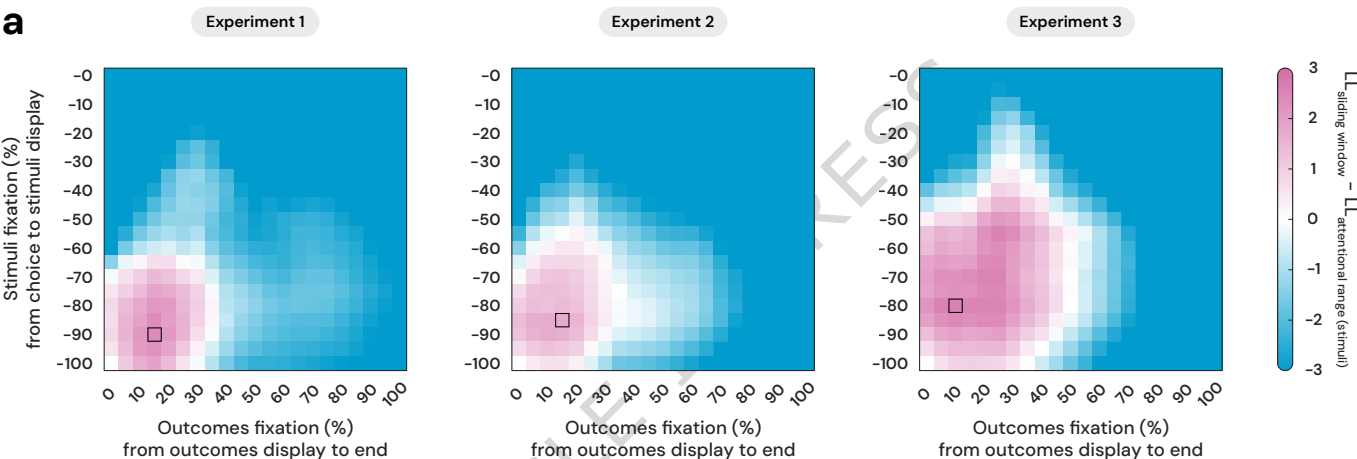


Experiment 3



a Experiment 2 (learning)**b** Experiment 3 (learning)**c** Experiment 2 choice phase**d** Experiment 3 outcome phase



a**b**