Accumulation of virtual tokens towards a jackpot reward enhances performance and value encoding in dorsal anterior cingulate cortex

Demetrio Ferro[1,2], Gabriela Mochol[1,2], Habiba Azab[3], Benjamin Hayden[3] and Rubén Moreno-Bote[1,2,4]

[1] Center for Brain and Cognition, Universitat Pompeu Fabra, 08002, Barcelona, Spain;

[2] Department of Information and Communications Technologies, Universitat Pompeu Fabra, 08002, Barcelona, Spain;

[4] Department of Neurosurgery, Baylor College of Medicine, Houston, TX, 77030;

[5] Serra Húnter Fellow Programme, Universitat Pompeu Fabra, Barcelona, 08002, Spain.

*Corresponding author: demetrio.ferro@upf.edu

Abstract

Normatively, our decisions ought to be made relative to our total wealth, but in practice, we make our decisions relative to variable, decision-time-specific set points. This predilection introduces a major behavior bias that is known as reference-point dependence in Prospect Theory, and that has close links to mental accounting. Here we examined neural activity in the dorsal anterior cingulate cortex (dACC) of macaques performing a token-based risky choice task, in which the acquisition of 6 tokens (accumulated over several trials) resulted in a jackpot reward. We find that subjects make faster and more accurate choices as the jackpot reward becomes more likely to be achieved, suboptimal behavior that can readily be explained by reference dependence. This biased behavior systematically covaries with the neural encoding of corresponding offer values. Moreover, we found significant enhancement in speed, accuracy and neural encoding strength for easier levels of difficulty in detecting the offer with the best expected value. These results suggest a neural basis of reference dependence biases in shaping decision-making behavior and highlight the critical role of value representations in dACC in driving those biases.

Introduction

In their seminal work, Kahneman and Tversky[1] fundamentally reshaped our understanding of decision-making under risk. While the utility and probability distortion curves aspect of Prospect Theory are important, reference dependence is just as important. Yet, while the neural basis of risk and probability distortion is well understood, the neural processes underlying reference dependence have yet to be delineated. In animal models, reference dependence can be difficult to study on single trials, but it can be done if we take the perspective of trial seriality in choice. Although previous works have studied the influence of behavioral history on decision-making[2–5], experimental paradigms typically ask subjects to choose between two probabilistic options followed by trial-based reward delivery. Yet, monkeys can understand token economies and can therefore be used to study the effects of endowment on choice and its neural representations.

Token-based decision-making tasks offer a nuanced approach to tracking behavior in cumulative sequences of decisions[15]. In these paradigms, choices result in the delivery of tokens and trial-by-trial decisions lead to token accumulation, which ultimately contributes to a final jackpot reward once a predetermined token count is reached. This framework allows researchers to explore how the perception of multi-trial reward probability, accumulated tokens, and proximity to jackpot achievement shape decision-making strategies and underlying neural processes.

Advances in the field of neurophysiology have substantially contributed to our understanding of neural mechanisms functionally involved in decision-making, implicating brain regions associated with reward processing such as the orbitofrontal cortex (OFC), the ventromedial prefrontal cortex (vmPFC), the anterior cingulate cortex (ACC), and the ventral striatum (VS)[6–14]. The vmPFC and OFC are predominantly associated with the evaluation of the relative value of alternative offers[15–21], while the ACC plays a role in monitoring conflicting information and actions[22–26], as well as the estimation of expected outcomes and the cognitive cost of the decision performance[27–30]. In particular, the dorsal ACC (dACC) has been linked to reward anticipation and cognitive effort computation[31,32], and to the processing of delayed rewards across various decision-making contexts[33–37]. Neural signals in ACC have been previously associated with multi-trial[38] and virtual reward expectation[39], with remarked impact on behavioral adjustments[40,41].

Using a token-based decision-making task (Fig. 1A)[21,33–35], we analyze data from two macaque subjects trained to choose between two alternative offers displayed on opposite sides of a visual display. The two offers contain visual cues of the associated risk/benefit level, quantified in terms of token count magnitude and probability. After the decision report, the outcome of the risky choice is revealed, and the associated positive or negative token count is added to the overall token count displayed at the bottom of the screen. The achievement of a net count of at least 6 tokens leads to the acquisition of a large jackpot reward.

We find that subjects use several factors to build their heuristic rules for trial-by-trial task performance. These include token count, as predicted by Prospect Theory. They also include factors that are not core elements of Prospect Theory, but that are consistent with the broader heuristics and biases paradigm, such as the difference in expected value (EV) of the two offers, and the mean value of the two offers, the risk level of the two offers (Fig. 2A). The model achieving best predictions consisted in the linear combination of the offer variables EV and risk. We observed that subjects are more accurate and faster for higher accumulated token counts, that is when they get closer to the acquisition of the jackpot reward (Fig. 2B-D). Combining these analyses with the difficulty in detecting the best offer, we found that larger accumulated token counts and lower difficulty in detecting the best option improve decision-making speed and accuracy (Fig. 3) and reduce the propensity for the riskiest option (Fig. 4). By using a linear model of the neural spike rate, we tracked the neural encoding of the SV of the two offers during task time, revealing that their encoding is significant and most pronounced during the time of the visual presentation of the respective offer (Fig. 5). Lastly, we found that larger accumulated tokens count at the start of the trial and easier detection of best offer SV both significantly enhance the neural encoding of SV in dACC, most prominently during offer presentation time (Fig. 6).

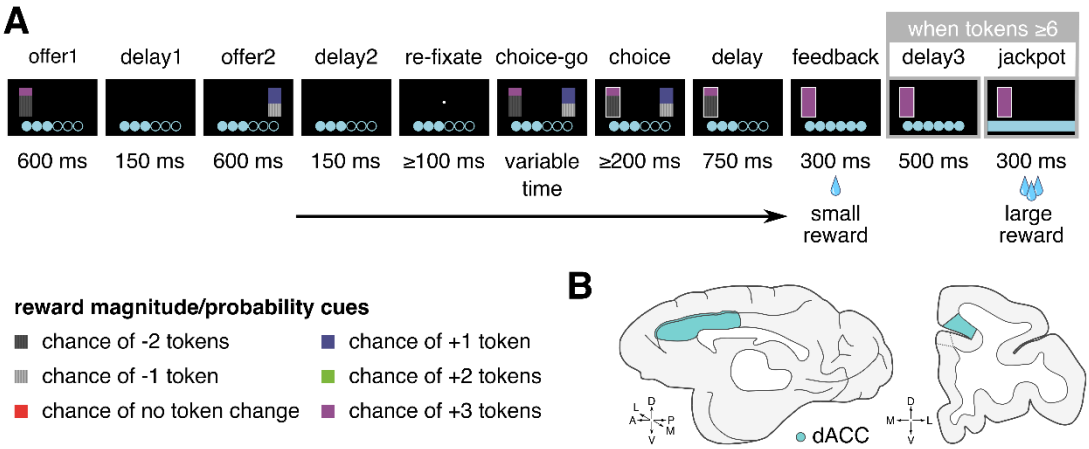

Figure 1. Behavioral task and recording sites. A. Two offers are sequentially presented (*offer 1-2*, 600 ms), interleaved by blank screen delays (*delay 1-2*, 150 ms). After the offers presentation, subjects are instructed to re-acquire fixation to a central cross (*re-fixate*) for at least 100 ms and report their choice after a *choice-go* cue, consisting of the presentation of both previous offer stimuli. Choice is reported via target fixation for at least 200 ms (*choice*). The outcome of the chosen probabilistic offer is drawn from uniform distributions, providing (positive or negative) virtual tokens upon risky choice resolution. A small fluid reward (100 μL) is provided in all trials. The token count is visible through execution time as initially unfilled circles at the bottom of the screen, filled by tokens as they are collected. At 6 tokens count, subjects receive a "jackpot", large reward (300 μL), and the count is reset. The height of the bar stimuli is informative of the associated reward probability, while the color is informative of the magnitude. The probability of the outcomes color-cued by the top and bottom parts of the stimuli are complementary. The probabilities are discretized as 10%, 30%, 50%, 70% and 90%, the magnitudes consist of reward counts (-2, -1, 0, +1, +2, +3) and could include negative values. Offers could also include safe options where 0 (red) or 1 (blue) token are achieved with 100% probability.

Results

The two subjects performed correct choices, i.e., choices that maximized outcomes, as measured by the offer expected value (EV), in most trials (subject 1: 79.78 ± 0.19% mean ± s.e.m., subject 2: 74.19 ± 0.18%). We found that choices were predominantly influenced by the difference in expected values (Δ_{EV}), the average magnitude of the two EVs (M_{EV}), the difference in risk between the offer with the best EV and the offer with worst EV (the offers risk level, ORL), and the tokens collected in task trials previous to current trial (accumulated tokens count, ATC) (Fig. 2A). These factors had a strong and statistically significant impact on the decisions made by the two subjects, suggesting that both relied on these variables to form heuristics and choose offers with the best EV.

Additionally, the presence of a jackpot in the previous trial (JPT) significantly influenced the choices of subject 1, whereas the outcome of the previous trial (OPT), i.e., the number of tokens collected in previous trial had a significant effect on the choices of subject 2. The number of trials since the last jackpot reward ($TSLR$) exhibited a moderate, though contrasting effect on the two subjects. Subject 1 made fewer choices for the offer with the best EV as the number of trials since the last jackpot reward increased, while subject 2 had the opposite tendency. These inter-subject differences are informative, but not central to our main investigations.

The role of accumulated tokens in choice performance

We found that the accumulated tokens count (ATC) has a strong impact on the subject's choices. When considering choices for the offer with the best offer EV, we found that subjects were more accurate (Fig. 2B) and faster (Fig. 2C) when they had collected more tokens, i.e. when they got closer to the possibility of achieving a jackpot reward. Conversely, the subjects take longer to perform the task and are less accurate when the accumulated tokens count is lower. To investigate this aspect further, we modeled the probability of choosing one of the two offers in relation to the difference in expected value of the two offers, by using a logistic model $\text{logit}(ch_1) = \beta_0 + \beta_1(EV_1 - EV_2)$, and fit the model separately in trials with 'Low' ($ATC = 0 - 1$) or 'High' ($ATC = 4 - 5$) accumulated tokens count (Fig. 2D). We found a significant increase in the slope of the logistic relationship ($\beta_{1,High} - \beta_{1,Low}$, p= $1.77 \cdot 10^{-4}$ in subject 1, p= $1.34 \cdot 10^{-19}$ in subject 2, two-tailed Wilcoxon signed rank test for zero median), corroborating previous insights about higher task engagement for higher accumulated tokens count, as shown by a significantly steeper logistic relationship (Fig. 2E).

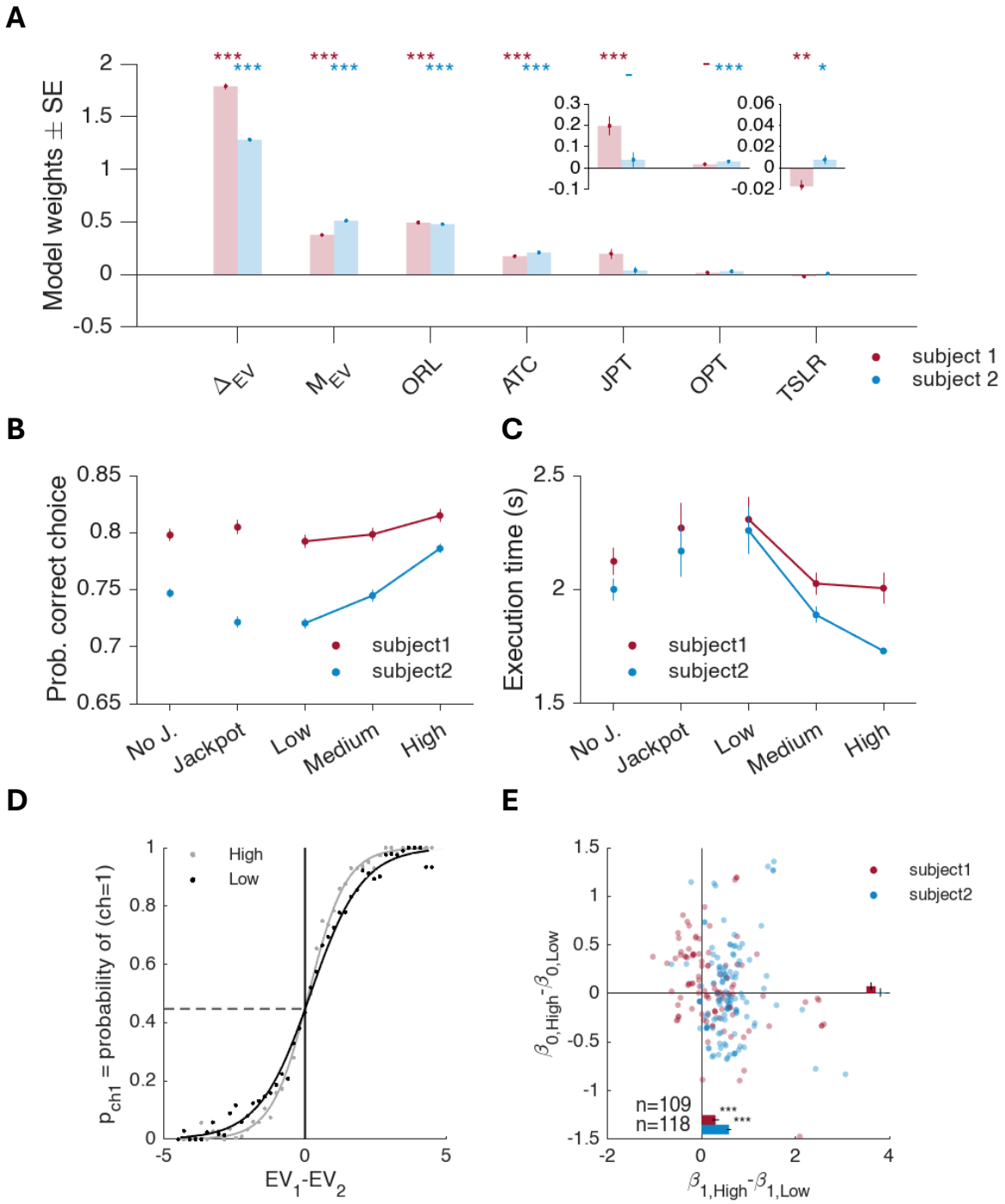

Figure 2. Value-based behavioral variables and accumulated tokens count influence decision-making performances.

A) Logistic model of the correct choice, i.e., choice for the offer with the best expected value (EV). We use $\Delta_{EV} = |EV_1 - EV_2|$ to define the discriminability of best EV, inversely related to task difficulty; $M_{EV} = (EV_1 + EV_2)/2$ the mean offer EV; the offer risk level (ORL) the difference between the risk of the offer with highest EV and the risk of the offer with lowest EV; the accumulated tokens count (ATC) as of the the start of current trial; the presence of a jackpot on previous trial ($JPT = 1$ jackpot, $JPT = 0$ no jackpot);

the number of trials since last jackpot reward ($TSLR$); and the outcome of previous trial (OPT), i.e. the number of tokens resulting from the risky choice made on previous trial. The analysis is applied separately for the two subjects (red, subject 1, $n = 47030$ trials; blue, subject 2, $n = 58911$, Supp. Table ST1). The intercept term $\beta_0 = -0.42 \pm 0.10$ in subject 1, and -0.60 ± 0.08 in subject 2, $p < 0.001$ in both subjects. **B)** Probability of correct choice vs accumulated reward at the start of each trial (Low: $ATC = [0, 1]$; Medium: $ATC = [2, 3]$; High: $ATC = [4, 5]$, Supp. Table ST1). **C)** Same as B but showing the task execution time. **D)** Logistic regression of first offer choice (ch1), designed as $\text{logit}(ch1) = \beta_0 + \beta_1(EV_1 - EV_2)$ for Low ($ATC = [0, 1]$) and High ($ATC = [4, 5]$) accumulated tokens count (data combined for the two subjects, High: $\beta_0 = -0.2^{***}, \beta_1 = 1.52^{***}, n = 24099$, Low: $\beta_0 = -0.21^{***}, \beta_1 = 1.09^{***}, n = 47563$, $^{***}p < 0.001$; Supp. Table ST1). **E)** Difference in (β_0, β_1) weights in D for High minus Low accumulated tokens count. Dots are weights differences in each session ($n = 109$, red for subject 1, $n = 118$, blue for subject 2). Bars show mean ± s.e.m. across sessions, significance is assessed via two-tailed Wilcoxon signed-rank tests, $p < 0.001$ for β_1 in both subjects.

Jackpot, accumulated tokens count, and best offer discriminability improve choice performances

We computed the fraction of errors in choosing the offer with the best EV based on whether the current trial was preceded or not by a jackpot reward, for different ranges of accumulated tokens count, and for different levels of difficulty in discriminating between the best and the worst offer. For accumulated tokens count, we used the ranges 'Low' ($ATC = [0, 1]$), 'Medium' ($ATC = [2, 3]$) and 'High' ($ATC = [4, 5]$) (Supp. Table ST1). For the difficulty, we used Δ_{EV} and the respective $\text{median}(\Delta_{EV}) = 1$ as difficulty discriminant, leading to split data in 'Hard' and 'Easy' trials, respectively based on whether $\Delta_{EV} < 1$ or $\Delta_{EV} \geq 1$ (Supp. Table ST2). Combining the jackpot conditions, accumulated tokens count, and difficulty, we find that the fraction of errors in choosing the offer with best EV most differentiate based on the trial difficulty (Fig. 3). The presence of jackpot before the current trial did not have a significant impact on the fraction of errors of subject 1 (Easy trials: No Jackpot 8.48±0.53% mean ± s.e.m.; Jackpot 8.03±0.62%, p= 0.067, two-tailed Wilcoxon signed-rank test; Hard trials: 31.10±0.53% No Jackpot; 30.18±0.79% Jackpot, p= 0.24), while in subject 2 the fraction of errors significantly increased after jackpot only in easy trials (Easy trials: 14.24±0.54% No Jackpot; 19.55±0.72% Jackpot, p= $1.21 \cdot 10^{-10}$; Hard trials: 35.95±0.44% No Jackpot; 35.46±0.68% Jackpot, p= 0.31, Fig. 3A).

The accumulated tokens count had a significant impact on the fraction of errors of both subjects, showing a decreasing trend in the fraction of errors in choosing the offer with the

best EV as the ATC increased (subject 1, Easy: 9.29±0.60% Low, 7.89±0.56% Medium, 7.28±0.53% High; subject 1, Hard: 31.58±0.60% Low, 31.36±0.76% Medium, 28.92±0.77% High; subject 2, Easy: 18.11±0.62% Low, 14.69±0.66% Medium, 9.25±0.44% High; subject 2, Hard: 37.25±0.49% Low, 35.65±0.66% Medium, 33.38±0.65% High, Fig. 3A). Note that the task trial occurrences become fewer and fewer as ATC values grow from Low to High, regardless of the task difficulty (Supp. Table ST1). Since the most prominent difference in fractions of errors was based on task difficulty, we also investigated fractions of errors for binned values of Δ_{EV}, which we used as an inverse metric of difficulty. As expected, the fraction of errors decreased with Δ_{EV}, as larger Δ_{EV} indicates easier detection of the best offer. By comparing the fractions of errors in High and Low accumulated tokens counts, we found a lower fraction of errors in High accumulated tokens, indicating that accumulating a larger number of tokens boosts the best EV discrimination, most prominent and significant in both subjects for intermediate difficulty (Δ_{EV}=[1, 3] subject 1, significant for all Δ_{EV} in subject 2, Fig. 3B). In similar way as for ATC, we find fewer trial occurrences when considering larger Δ_{EV} bins, coinciding with easier trials (Supp. Table ST2). These results align with the observation that subjects tend to be more accurate and faster whenever they accumulate more tokens before the current trial, adding a further level of stratification in considering the difficulty, that has a prominent role in committing best offer contingency detection errors.

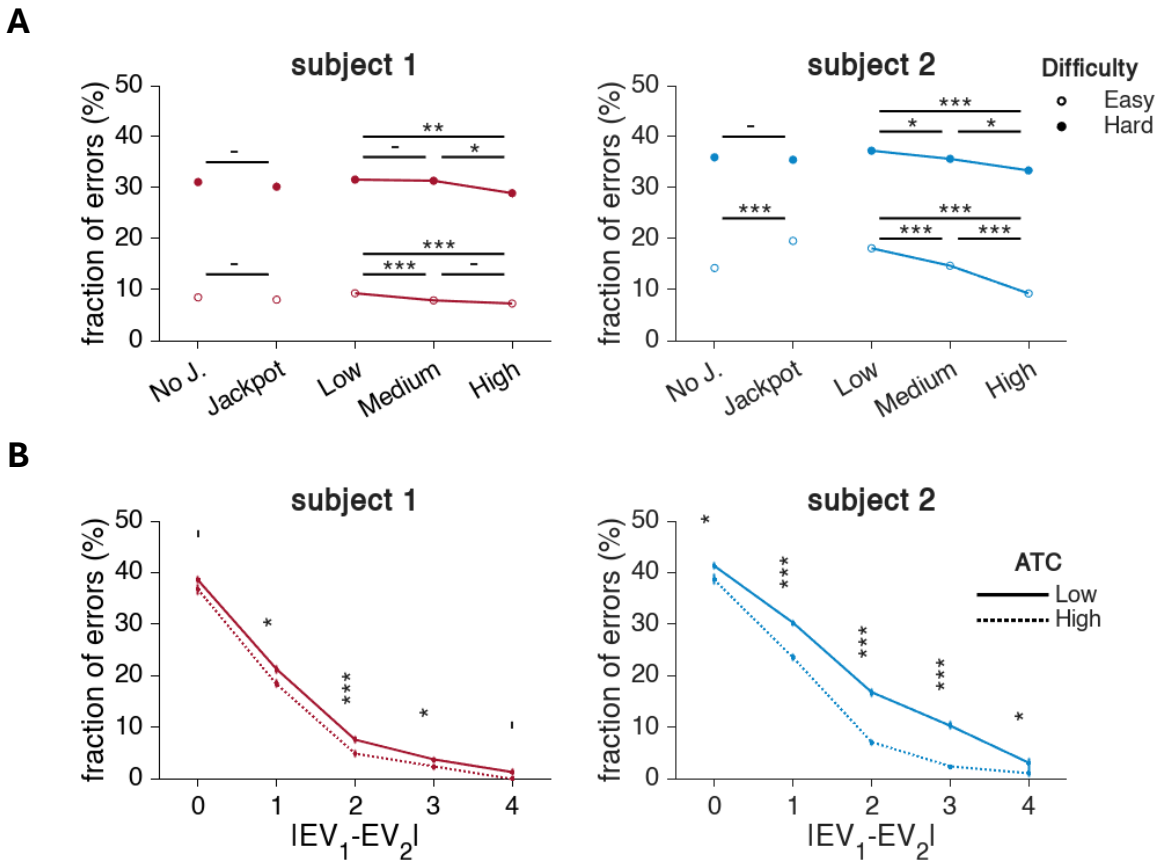

Figure 3. Fraction of errors in reporting correct choice decreases with accumulated tokens count and task difficulty. A) Average fraction of errors (subject 1, n=109 sessions; subject 2, n=118) for cases of no jackpot ($JPT = 0$) or jackpot ($JPT = 1$) on previous trial and for different ranges of accumulated tokens count as of the start of current trial (Low: $ATC = [0, 1]$; Medium: $ATC = [2, 3]$; High: $ATC = [4, 5]$, Supp. Table ST1). Data are split based on difficulty as 'Hard' (filled markers, $\Delta_{EV} <$ median(Δ_{EV})) and 'Easy' (empty markers, $\Delta_{EV} \geq$ median(Δ_{EV})) (Supp. Table ST2). Left: subject 1, right: subject 2. Differences in ATC ranges are tested via Wilcoxon signed-rank tests, FDR corrected via Benjamini-Hochberg procedure (*p< .05, **p< .01, ***p< .001). **B)** Average fraction of errors (subject 1, n=109; subject 2, n=118) for binned values of difficulty, inversely related to $\Delta_{EV} = |EV_1 - EV_2|$ (Supp. Table ST2). Data are split in 'Low' (solid, $ATC = [0, 1]$) and 'High' (dotted, $ATC = [4, 5]$) accumulated tokens count. Left: subject 1, right: subject 2. Differences in bins are assessed via two-tailed Wilcoxon rank sum (*p< .05, **p< .01, ***p< .001).

Jackpot, accumulated tokens count, and difficulty in best offer discriminability impact risk propension

We asked whether these results could relate to a higher propensity for risky options whenever subjects feel that achieving a jackpot reward is less likely and far apart, requiring further

accumulation of tokens. By comparing the fraction of trials where subjects chose the offer with the highest risk, we observed that the presence of a jackpot on the previous trial decreased the fraction of choices for the riskiest option, significant for subject 1 (subject 1: No Jackpot 59±0.43% mean ± s.e.m., Jackpot 58±0.71%; p= $1.04 \cdot 10^{-3}$, two-tailed Wilcoxon signed-rank test, subject 2: No Jackpot 62.27±0.36%, Jackpot 62.04±0.59%; p= 0.37). We found that both subjects significantly made more risky choices when choosing the offer with best EV was hard ($\Delta_{EV} < 1$) as opposed to easy ($\Delta_{EV} \geq 1$) trials (subject 1: Easy 50.65±0.48% mean ± s.e.m., Hard 68.25±0.53%, p= $1.47 \cdot 10^{-19}$, two-tailed Wilcoxon signed-rank test, subject 2: Easy 54.69±0.41%, Hard 69.64±0.44%, p= $4.21 \cdot 10^{-21}$).

By combining difficulty and jackpot conditions, we found that subjects consistently chose risky offers less frequently after a jackpot reward in Hard trials (subject 1: 'No Jackpot' 68.83±0.33%, 'Jackpot' 66.66±0.73%; subject 2: 'No Jackpot' 70.24±0.29%, 'Jackpot' 67.76±0.66%; Fig.4A), while in Easy trials risky choices after jackpot were less frequent in subject 1 ('No Jackpot' 50.97±0.37%, 'Jackpot' 49.38±0.73%), and more frequent in subject 2 ('No Jackpot' 54.50±0.32%, 'Jackpot' 56.40±0.71%; Fig. 4A). We reported a remarked difference in the fraction of choices for the riskiest option between Easy and Hard trials, with Hard trials involving higher fractions of choices for the riskiest option across ATC conditions (subject 1: 17.68±0.79% mean ± s.e.m.; subject 2: 15.70±0.78%; Fig. 4A). We found a consistent inverted U-shape relationship between the fraction of risky choices and ATC ranges in Hard trials in both subjects, showing increasing trend for $ATC = [0, 2]$, and decreasing trend for $ATC = [3, 5]$ (Fig. 4A; subject 1: Low 68.14±0.45%, Medium 69.39±0.52%, High 67.68±0.63%; subject 2: Low 69.35±0.40%, Medium 72.36±0.46%, High 67.01±0.59%). To further investigate the effect of risk on the decisions of the two subjects, we quantified the risk attitude of subjects by using a Markowitz model, i.e., computing the ratio between the weight of the risk (β_2) and EV (β_1) differences between the two offers in a logistic model of the choice including these two variables as choice predictors (Methods 2.4, accuracy as in the model m_2 in Supp. Fig. S2). The model parameter $\theta = -\beta_2/\beta_1$ describes the subjective behavioral attitude as risk aversion ($\theta > 0$) or risk-seeking ($\theta < 0$). We report a risk-seeking attitude in both subjects, with higher risk aversion (reduced magnitude of θ) as the number of accumulated tokens increases, and for Easy trials (Fig. 4B).

Comparing jackpot conditions, we find that θ has smaller, negative magnitude for Hard trials following a jackpot reward, consistent in the two subjects (Fig. 4B; subject 1: -0.25 No-Jackpot mean, -0.21 Jackpot, subject 2: -0.33 No Jackpot, -0.30 Jackpot), in line with lower risky choice fractions results (Fig. 4A).

We found that the EV difference has an increasing impact (β_1) on choices for the riskiest option, consistent in the two subjects for Hard trials (Fig. 4C, subject 1: 2.17 ± 0.08 Low mean ± s.e.m, 2.41 ± 0.10 Medium, 2.47 ± 0.12 High, subject 2: 1.43 ± 0.07 Low, 1.81 ± 0.09 Medium, 1.93 ± 0.10 High) than in Easy trials (Fig. 4C, subject 1: 1.62 ± 0.05 Low, 1.74 ± 0.06 Medium, 1.79 ± 0.07 High, subject 2: 1.07 ± 0.03 Low, 1.38 ± 0.04 Medium, 1.68 ± 0.06 High). The risk difference weight (β_2) shows an inversed U-shape, most prominent and consistent in the two subjects for Hard trials (Fig. 4D, subject 1: 0.52 ± 0.02 Low mean ± s.e.m, 0.61 ± 0.03 Medium, 0.55 ± 0.03 High, subject 2: 0.49 ± 0.02 Low, 0.66 ± 0.02 Medium, 0.47 ± 0.02 High), larger than for Easy trials (Fig. 4D, subject1: 0.39 ± 0.03 Low, 0.38 ± 0.03 Medium, 0.34 ± 0.04 High, subject 2: 0.40 ± 0.02 Low, 0.54 ± 0.03 Medium, 0.39 ± 0.03 High), both results are in line with the fraction of risky choices (Fig. 4A). The Markowitz variable $\theta = -\beta_2/\beta_1$, shows decreased magnitude as the accumulated token counts increases for 'High' ATC ranges, in both Hard (Fig. 4D; subject 1: -0.24 Low mean, -0.25 Medium, -0.22 High; subject 2: -0.34 Low, -0.37 Medium, -0.24 High) and Easy trials (Fig. 4D; subject 1: -0.24 Low mean, -0.22 Medium, -0.19 High; subject 2: -0.38 Low, -0.39 Medium, -0.23 High), suggesting that when the jackpot reward becomes more likely in tokens count, the subjects are more averse to risk. All the above results found for Low, Medium, High ATC ranges are also found for discrete values of ATC (Supp. Fig. S1).

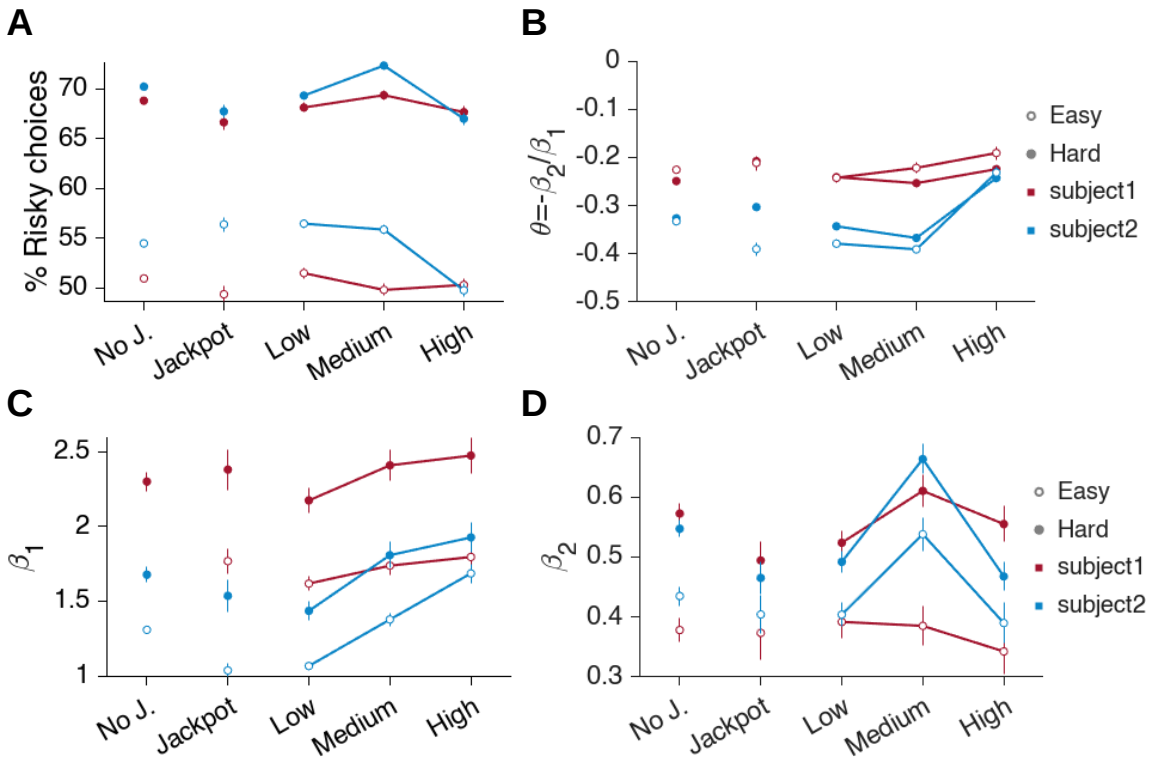

Figure 4. Risk seeking attitude increases with accumulated tokens count and easier best offer discriminability. A) Fraction of trials (mean ± s.e.m.) with choice for the offer with higher risk. The data are split in 'No Jackpot' ($JPT = 0$), 'Jackpot' ($JPT = 1$), and for accumulated tokens count ('Low' $ATC = [0, 1]$, 'Medium' $ATC = [2, 3]$, 'High' $ATC = [4, 5]$). Data are split for subject 1 (red) and subject 2 (blue), and for Easy ($\Delta_{EV} \geq 1$, empty markers) and Hard ($\Delta_{EV} < 1$, filled markers). **B)** Markowitz risk return model for the offer utility based on the mean value (EV) and risk (R) of the offers. The model parameter (θ) describes risk attitude ($\theta < 0$ risk seeking, $\theta > 0$ risk avoiding) for jackpot cases and for values of accumulated token counts. Data is split in difficulty and across subjects as in A. **C)** Markowitz risk return model, parameter β_1 relative to EV weights for jackpot cases and for values of accumulated token counts. Data is split as in A. **D)** Markowitz risk return model, parameter β_2 relative to risk R weights for jackpot cases and for values of accumulated token counts. Data is split as in A.

Neural encoding of Subjective Value: tuning and activation in presentation epochs

In all neural analyses we summarized the value of each offer in a single variable, the subjective value (SV), defined by a linear combination of EV and risk R. The SV definition followed the test of multiple models for the choices of the two subjects, leading to the selection of the model with the best choice prediction accuracy (Methods 2.4, Supp. Fig. S2). We detected the encoding of the SV of the two offers by using a time-resolved assessment of a linear model of the firing rate (Methods 3), including the two SVs, and fit independently for each of the recorded cells. We found cells with positive (Fig. 5A, orange, $n = 34$ cells, sessions and subjects combined) or negative (Fig 5A, purple, $n = 41$ cells) significant encoding of SV_1. The time-resolved fraction of total cells showing positive or negative encoding tuning is comparable (Fig. 5B, orange and purple), with prominent significance soon after the first offer is presented and through the second offer presentation, gradually decaying over time. Similarly, for the second offer, we found a comparable number of cells showing positive (Fig. 5C, $n = 53$ cells) or negative (Fig. 5C, $n = 48$ cells), with the time-resolved fractions of significant cells showing more prominent significance after offer presentation, followed by gradual decay.

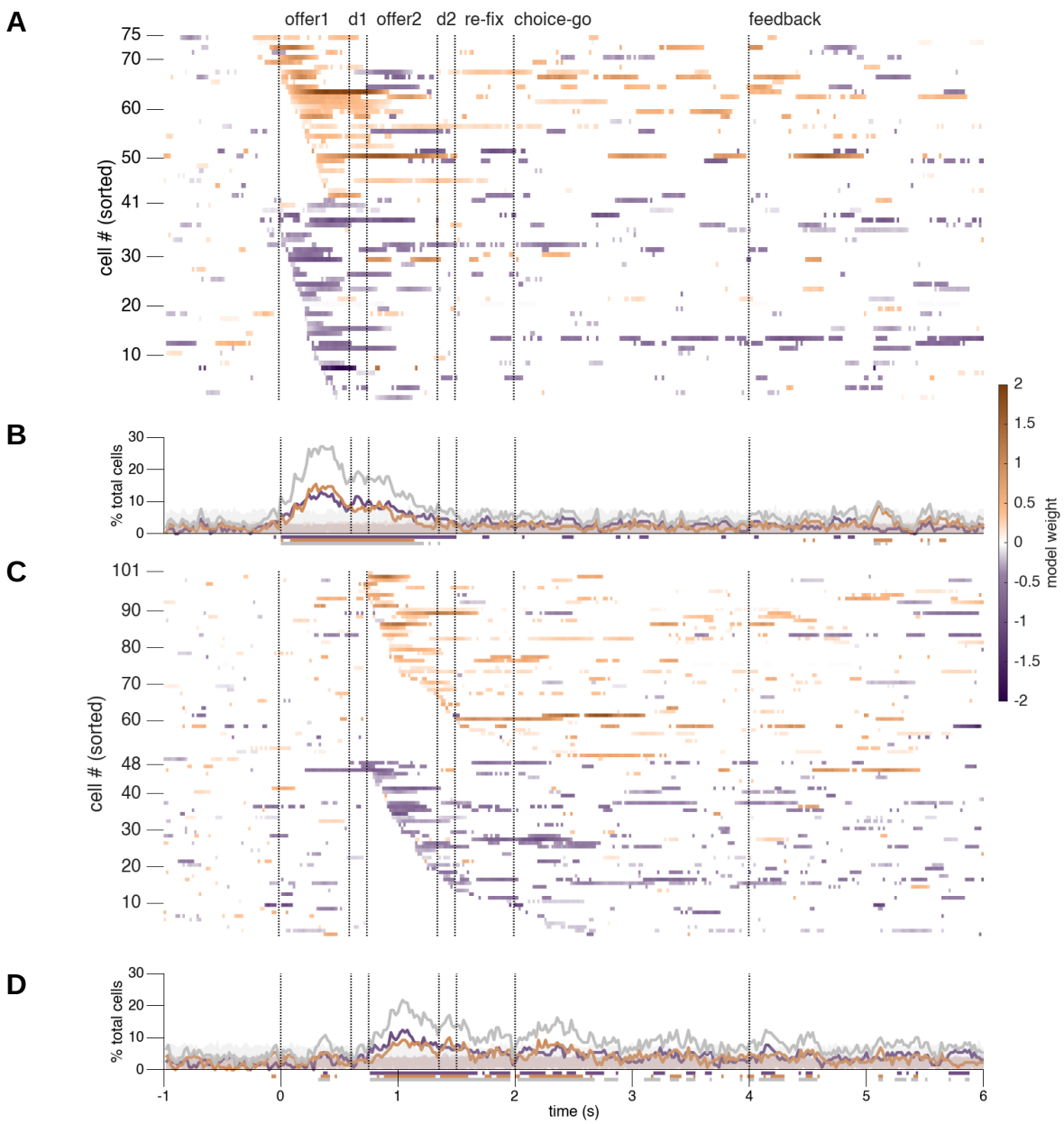

Figure 5. Neural encoding and tuning of SVs is most prominent during offer presentation time. A. Encoding of SV_1 across recorded cells (y-axis) during task time (x-axis). Significant bins are color-coded by model weight sign (β_1; orange: positive, purple: negative) and intensity (darker for larger β_1 magnitudes). Cells are sorted by the time of the first significant encoding run across consecutive time bins. Task epochs labels d1: delay1, d2: delay2. **B.** Fraction of total cells significantly encoding SV_1 (y-axis) during task time (x-axis), gray for significant (both positive and negative) weights, or color-coded by model weight sign as in A. Shaded areas show the 95th percentile of the same fractions for trial-order shuffled data. Solid lines at the bottom of the panel indicate significant runs of consecutive time bins where empirical fractions exceed the 95th percentile of shuffled data, further assessed in length via run-length analyses. **C-D.** Same as A-B, but for SV_2.

Accumulated tokens count and best offer discriminability enhance the neural encoding of SV

We investigated the fractions of cells showing significant encoding of the two SVs for different ranges of ATC ('Low' $ATC = [0, 1]$, 'High' $ATC = [2, 5]$, Methods 3.1), and for different levels of difficulty in best offer EV discriminability ('Easy' $\Delta_{SV} \geq median(SV)$, 'Hard' $\Delta_{SV} < median(SV)$, Methods 3.2), by also balancing ATC and difficulty trial pools sizes prior to neural encoding assessments. We observe that the fractions of cells with significant encoding of SV_1 show increases above the significance threshold soon after the first offer is presented and through the second offer presentation, including the respective delays (Fig. 6A). Interestingly, for High ATC we find smooth, gradual transition (Fig. 6A, dark green), while for 'Low' ATC we found a deflection towards the end of the first offer presentation, followed by a delayed increase around first delay and second offer onset (Fig. 6A, light green). This coincided with significantly larger fractions of cells for 'High' ATC in delay 1 (Fig. 6E top, p= $3.24 \cdot 10^{-3}$, one-tailed Wilcoxon signed-rank test), offer 2 (Fig. 6E top, p= $3.60 \cdot 10^{-2}$) and delay 2 (Fig. 6E top, p= $2.20 \cdot 10^{-3}$). For SV_2, we found that the fractions of significant cells showed a prominent significant increase after the second offer presentation, gradually decaying through the following task epoch (Fig. 6B), with larger fractions for 'High' ATC in offer 2 (Fig. 6E bottom, p= $2.45 \cdot 10^{-2}$, one-tailed Wilcoxon signed-rank test). By computing the fractions of significant cells across difficulty levels, we found that SV_1 is significantly encoded after the first offer presentation (Fig. 6C), significantly more prominent in Easy trials in offer 1 (Fig. 6F top, p= $4.72 \cdot 10^{-3}$, one-tailed Wilcoxon signed-rank test), delay 1 (Fig. 6F top, p=$1.60 \cdot 10^{-2}$) and offer 2 (Fig. 6F top, p= $4.57 \cdot 10^{-3}$). Similarly, for SV_2 we found significant fractions of encoding cells after the second offer presentation (Fig. 6C), significantly higher in Easy trials at offer 2 epoch time (Fig. 6F bottom, p= $8.16 \cdot 10^{-3}$, one-tailed Wilcoxon signed-rank test).

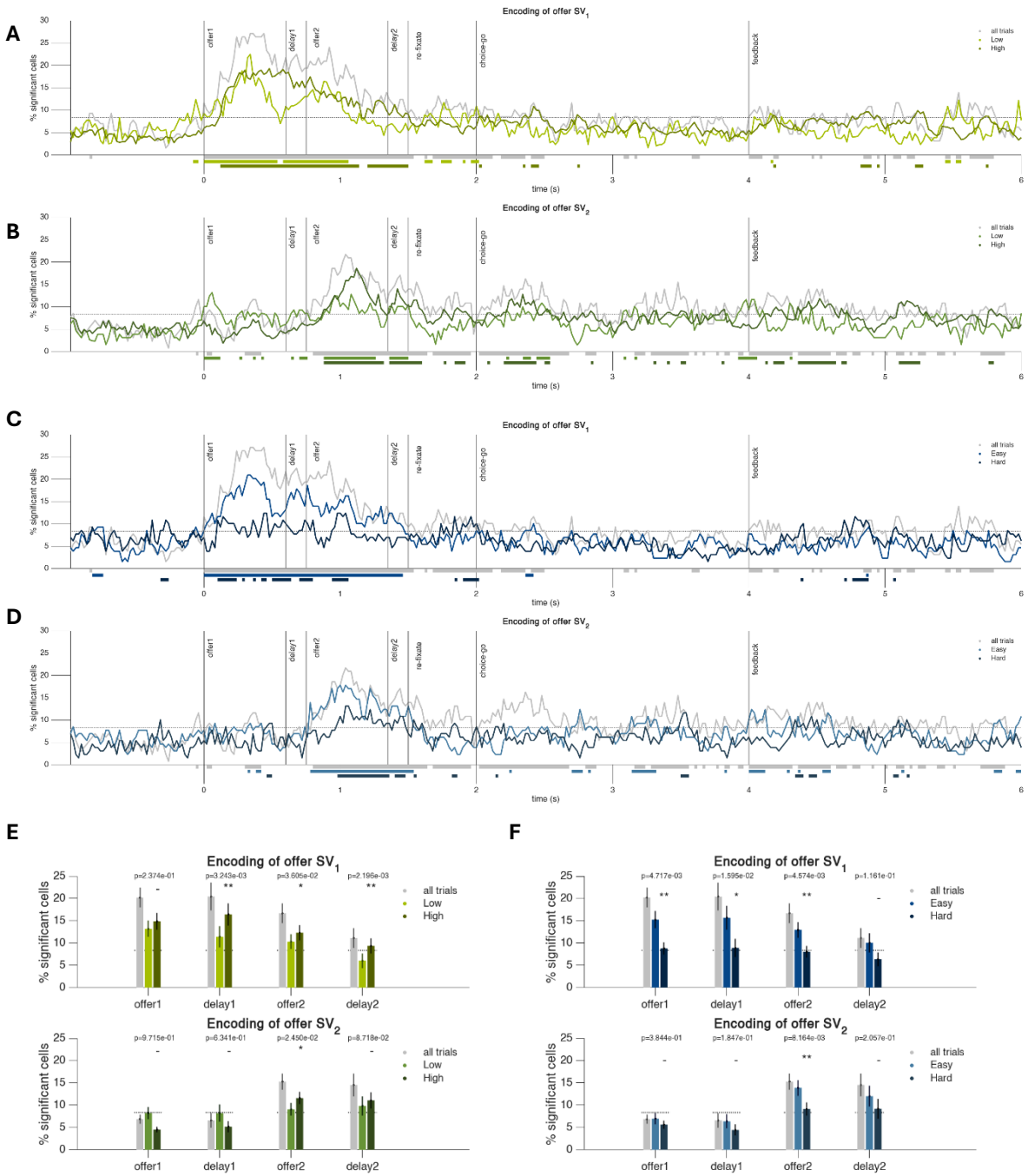

Figure 6. Neural encoding of SV is enhanced for higher tokens count and for easier best offer discriminability.

A. Fraction of significant cells encoding SV_1 for 'Low' (light green, $ATC = [0, 1]$ tokens) and 'High' (dark green, $ATC = [2, 5]$ tokens) accumulated reward from previous trials, at the start of the trial. In gray, the fraction of cells encoding SV_1 across all trials is. Since the two subsets 'Low' and 'High' have slightly different size ($\cong 45.18\%$ of total trials has 'Low' ATC and $\cong 54.82\%$ has 'High' ATC), the analysis is repeated over $n = 10$ randomly assigned subsets of 'High' ATC trials of size equal to the size of the 'Low'

ATC trials set ($\cong 45.18\%$ of total trials). The results shown are averages across the $n = 10$ randomly assigned resampled subsets. The dotted line shows the 95th percentile of the same fractions run over shuffles of the trial order. The solid lines at the bottom show runs of consecutive bins with significant fractions, exceeding the 95th percentile (dotted line), further assessed in length via run-length analyses. **B.** Same as C, but for SV_2. **C.** Fraction of significant cells encoding SV_1 for 'Easy' (light blue, $\Delta_{SV} >$ median(Δ_{SV})), 'Hard' (dark blue, $\Delta_{SV} \leq median(\Delta_{SV})$), and all trials (gray). Using the median as discriminant allows to split the trials in equally sized 'Easy' and 'Hard' subsets. **D.** Same as A, but for SV_2. **E.** Epoch-wise assessment of results in A-B. The fractions of time bins showing significant encoding in each cell are computed in main task epochs (*offer 1*, *delay 1*, *offer 2*, *delay 2*). Cell-wise averages are compared via one-tailed signed-rank tests between difficulty conditions, assessing that High ATC trials show larger significant encoding fractions than Low ATC trials. As in A-B, results are resampled in equally sized subsets for High and Low ATC number of trials. Resampled results are averaged across the $n = 10$ subsets prior to statistical assessment. **F.** Same as E but for results in C-D, assessing that Easy trials have larger significant encoding.

Discussion

This study shows that the accumulation of multi-trial tokens to achieve a final jackpot has multiple implications on trial-based decision-making, both in behavior and at the neural level. The accumulation of virtual tokens towards a jackpot reward has important effects on the accuracy and on the speed of subjects solving the task. The behavioral choices were faster and more accurate for higher accumulated tokens, suggesting enhanced efficiency in the computation of task contingencies whenever the jackpot reward achievement was more likely, corresponding to improved accuracy in choosing the most valuable offer. These effects coincided with an improved detection of best offer contingencies, as indicated by negative trends in the fractions of errors in choosing the best offer for higher accumulated tokens counts, and to a lower propensity for risk-taking whenever the cumulative jackpot started to become more likely. Importantly, both results showed significant differences based on task difficulty, with easier trials being involved in much lower levels of errors and risky choices. At the neural level, we found that the encoding of the value of the two offers was significantly prominent at their respective presentation time, with gradual fading throughout subsequent task epochs. Finally, we found that higher accumulated tokens count, and easier task difficulty enhance the strength of the neural encoding of value in dACC cells, revealing the importance of considering cumulative reward levels in neural analyses during task

execution. These results constitute an important novelty factor as studies in the field are often not emerging or only provide limited evidence for multi-trial reward factors[38] and for virtual rewards[39].

While results align with previous knowledge about behavioral theories for subjective preferences[1,42,43], studying the effects of token-based accumulation on trial-based decisions showed its relevant impact on choice accuracy and performance time at discrete cumulative stages before reward attainment, revealing relevant influence on best choice discrimination and risk-taking attitude. Importantly, the use of positive or negative tokens led to formulate the reinforcement of virtual rewards and punishments, in turn, related to risk-taking and learning, most often studied at a coarser level via functional magnetic resonance imaging[44,45].

Converging evidence suggests that expected rewards are updated by the animal experience and that the estimation of such expected rewards occurs in the frontal cortex and basal ganglia[46]. Our results showing enhanced neural encoding for evaluations close to reward achievement, and with harder discriminability of best offer choice can be linked to previous results suggesting that more accurate economic evaluations are associated with increased activity in the prefrontal cortex[44] with stronger activation linked to the recruitment of more complex decision-making, with relevant impact on the assessment of potential economic gains and losses[10]. When rewards are readily available, neurons have been found to show enhanced activation, favoring choices toward immediate achievement[16]. This might have notable implications, connecting the task inverse U-shaped risk-taking performances to increases in dopamine release in neural reward circuits[47], which can drive to increased motivation and goal-directed behavior, regulating decision in human reinforcement learning[48], linking the prioritization of decisions that maximize exposure to rewards[49–51] to the enhancement of their neural encoding.

Furthermore, the involvement of the dACC in token accumulation reinforcement is particularly noteworthy. Previous research has implicated the dACC in cognitive control[7,52], conflict monitoring[22–24], and value-based decision-making[12,14,25,36]. Our findings further support its role in encoding the subjective value of outcomes in a dynamic, multi-trial context. The enhancement in neural encoding of the subjective choice value in the dACC as the trials became closer to jackpot attainment suggests that this brain region not only

processes immediate choices but also tracks accumulated reward across time, updating the decision-making strategy as the reward outcome becomes increasingly imminent. This suggests that subjects monitor a multi-trial probabilistic notion of jackpot achievement likelihood, which is often overlooked when investigating the putative functional role of value encoding in dACC.

The dynamic tracking of tokens accumulation towards a jackpot reward has the potential to inform the development of more realistic decision-making models that consider the history of virtual rewards, or reinforcement-based reward probability evolution, expanding current knowledge about multi-trial history shapes subsequent choices[2–5,53,54]. In conclusion, this study underscores the importance of considering accumulated rewards in decision-making paradigms, both from a behavioral and neural perspective. By dynamically tracking neural value encoding and incorporating multi-trial factors into our understanding of decision-making, we gain a more comprehensive view of the neural functions underlying the adaptive processes that guide decision strategies.

Methods

1. Experimental settings and reproducibility

The data includes $n = 227$ behavioral sessions from two subjects (109 sessions, 433.28 ± 8.70 mean ± s.e.m. trials per session for subject 1, 118 sessions, 500.68 ± 10.97 trials for subject 2), of which $n = 108$ sessions (65 sessions, 479.16 ± 14.57 mean ± s.e.m. trials per session for subject 1, 43 sessions, 519.35 ± 14.62 trials for subject 2) include extracellular activity of dACC (Area 24) recorded using single electrodes (Frederick Haer & Co., impedance range 0.8 ± 4 MΩ), while monkeys performed the task. The electrode probes were lowered using a microdrive (NAN Instruments) until well-isolated activity was encountered. Data covered 1-3 simultaneous single neurons, Plexon system (Plexon, Inc.) was used to isolate the action potential waveforms. Cells are included based on quality of isolation but not on task–related response properties. Here we present neuronal data consisting of $n = 129$ single units (55 for subject 1 and 74 for subject 2). These data were published in the context of different research questions[34,35].

1.1 Experimental Procedures and Neural Recordings

All procedures were approved by the University Committee on Animal Resources at the University of Rochester and were designed and conducted in compliance with the Guide for the Care and Use of Animals of the Public Health Service's Guide for the Care and Use of Animals (protocol UCAR-2010-169). Two macaques (*Macaca mulatta*) served as subjects. Recorded brain regions covered dACC, using single-contact probes. Data include 47228 behavioral trials (433.28 ± 8.70, mean $\pm$ sem across sessions) from subject 1 and 59080 (500.68 ± 10.97) from subject 2. Of these, 20604 (479.16 ± 14.57) from subject 1, and 33758 (519.35 ± 14.62) from subject 2 were simultaneous to dACC data recordings.

1.2 Behavioral Task

The token-based reward gambling task consists of the sequential presentation of two visual stimuli at the opposite sides of the screen, providing reward offers to be subsequently chosen by performing saccade to either target location (Fig. 1A)[21,33–35]. The task starts with a first offer presentation (*offer 1*, 600 ms), followed by a first delay time (*delay 1*, 150 ms), a second offer presentation (*offer 2*, 600 ms), respectively followed by a second delay time (*delay 2*, 150 ms). After fixating a central cross (*re-fixate*), for at least 100 ms, the choice can be reported upon *choice-go* cue onset, consisting of the simultaneous presentation of both offer stimuli previously presented. Choice is reported by fixation on target offer cue for at least 200 ms (*choice*). A successful choice selection fixation was followed by 750 ms delay period after which the gamble was resolved. Regardless of the outcome, within the next 300 ms a small juice reward (100 μL) was delivered to keep subjects engaged in the task. If, after the outcome resolution less than six tokens are collected, the trial ends. If six or more tokens are collected, after a 500 ms delay period a “jackpot” large liquid reward (300 μL) is delivered, the token count is reset, and the trial ends. Trials were separated by a random inter–trial interval ranging between 0.5 and 1.5 s. Circular indicators at the bottom of the screen report the token count through task time as empty circles, filled according to token count. The visual offer stimuli are split horizontally into top and bottom parts of different colors, with associated reward magnitude and probability. The height of the top/bottom part is informative of the associated reward probability, while the color is informative of the magnitude. Thanks to the split, the probability of the outcomes color-cued by the top and bottom parts of the

stimuli is always complementary. The probabilities are discretized as $10\%, 30\%, 50\%, 70\%$ and 90%, the magnitudes consist of reward counts $(-2, -1, 0, +1, +2, +3)$ including negative values. In addition, offers include safe options where 0 (red) or 1 (blue) token are achieved with 100% probability. Each gamble includes at least one positive or zero outcome. This allows a less trivial level of choice computation and keeps the subjects motivated throughout behavioral execution.

2. Behavioral Data Analyses

We analyzed behavioral data at multiple levels. First, we assessed the factors that influenced the subjects' choices. We defined the expected value by considering the magnitude v and probability p of the top (t) and bottom (b) parts of the offers, as: EV $= p^t v^t + p^b v^b$, where $p^b = 1 - p^t$. Similarly, we defined the risk associated with the offers as: $R = p^t(v^t - EV)^2 + (1 - p^t)(v^b - EV)^2$. In behavioral data analyses, we stratified the analyses in ranges of accumulated tokens: 'Low', 'Medium', 'High, and in levels of difficulty in detecting the offer with best EV: 'Easy' and 'Hard'. As accumulated tokens count ranges, we use 'Low' $ATC = [0, 1]$, 'Medium' $ATC = [2, 3]$, and 'High' $ATC = [4, 5]$ (Supp. Table ST1).We defined difficulty as inversely related to the absolute EV difference $\Delta_{EV} = |EV_1 - EV_2|$, i.e., larger Δ_{EV} indicate lower difficulty. We used the median value of Δ_{EV}, $median(\Delta_{EV}) = 1$ as discriminant for 'Easy' ($\Delta_{EV} \geq 1$) and 'Hard' ($\Delta_{EV} < 1$) trials (Supp. Table ST2).

2.1 Logistic model of choice

To assess the choice performance, we used a logistic model to classify the binary choice for either offer. Other than the EV difference between the two offers (Δ_{EV}), we also used as regressors : the 'Mean expected value' as the average EV in each trial: $M_{EV} = (EV_1 + EV_2)/2$; the 'Offer Risk Level' $ORL = R_{Best} - R_{Worst}$ where R_{Best} or R_{Worst} indicates the risk associated to the offer with best or worst EV, respectively; the 'Accumulated Tokens Count' ATC, i.e., the token count at the start of the trial; the 'Trials Since Last Reward' $TSLR$ that is, the count, in number of trials, since last jackpot reward; the 'Outcome of Previous Trial' OPT , that is the number of tokens achieved in previous trial; and the 'Jackpot on Previous Trial' JPT, a binary variable $= 1$ if previous trial ended with jackpot, $= 0$ otherwise. Due to using previous trial variables, the first trial in each session was removed in this analysis ($n = 47030$ total trials in subject 1, $n = 58911$ total trials in subject 2). The logistic model is designed to model the choice of 'correct choice' CC, in turn defined as choice for the offer with best EV. In symbols, $CC = 1$ if choice is for offer 1 when $EV_1 > EV_2$, and for offer 2 when $EV_2 > EV_1$; $CC = 0$ in all other cases. The logistic model assumes a logistic relationship between the correct choice CC and the above variables: $\text{logit}(CC) = \beta_0 + \beta_1 \Delta_{EV} + \beta_2 M_{EV} + \beta_3 ORL + \beta_4 ATC + \beta_5 TSLR + \beta_6 OPT + \beta_7 JPT$.
Prior to least squares estimation of the weights $\beta_{i=[0,7]}$, the regressed variables are normalized to their absolute maximum value so that (Δ_{EV}, ATC, $TSLR, JPT$) are in the range [0, 1], and (M_{EV}, ORL, OPT) are in the rage [-1, 1], and assessed for significance via two-tailed F-Statistics tests (Supplementary Materials).

2.2 Behavioral choice accuracy and execution time

In addition to the choice models described above, we aim to investigate the effects of accumulated reward. In Fig. 2B we investigate the fraction of trials with correct choice (i.e.,

choice for the offer with best EV) for '*No Jackpot*' (JPT $= 0$) or '*Jackpot*' (JPT $= 1$) on previous trial, for 'Low' $ATC = [0, 1]$, 'Medium' $ATC = [2, 3]$, and 'High' $ATC = [4, 5]$ accumulated tokens count (Supp. Table ST1). The same conditions are used to analyze the fraction of errors (i.e., trials where the subjects did not choose the offer with best EV), as well as for binned levels of Δ_{EV}, inversely related to the difficulty in detecting the best offer (Supp. Table ST2).

2.3 Risk propension

The Markowitz risk return model is defined via the weights of the logistic regression of offer choice (*ch1*$= 1$ for first offer choice, 0 otherwise) as $\text{logit}(ch1) = \beta_0 + \beta_1(EV_1 - EV_2) + \beta_2(R_1 - R_2)$. The subject's utility for the ith offer is modelled as a trade-off between EV and risk, defined as $U_i = EV_i - \theta R_i$. The model parameter $\theta = -\beta_2/\beta_1$, describes the behavioral attitude as risk aversion ($\theta > 0$) or risk seeking ($\theta < 0$).

2.4 Subjective values

The definition of subjective value followed the assessment of alternative models of choice prediction, including task variables such as value of the top and bottom part of the offers (v^t, v^b), expected values (EV_1, EV_2), risks (R_1, R_2). We defined the models:

m_0: $\text{logit}(ch1) = \beta_0 + \beta_1 v_1^t + \beta_2 v_1^b - \beta_3 v_2^t - \beta_4 v_2^b$;

m_1: $\text{logit}(ch1) = \beta_0 + \beta_1 EV_1 - \beta_2 EV_2$;

m_2: $\text{logit}(ch1) = \beta_0 + \beta_1 EV_1 - \beta_2 EV_2 + \beta_3 R_1 - \beta_4 R_2$;

m_3: $\text{logit}(ch1) = \beta_0 + \beta_1 v_1^t + \beta_2 v_1^b - \beta_3 v_2^t - \beta_4 v_2^b + \beta_5 EV_1 - \beta_6 EV_2 + \beta_7 R_1 - \beta_8 R_2$.

The model weights were fit via ordinary least squares regression and the accuracy in predicting the subject's choice was computed in all models, cross-validating over $k = 4$ data folds. The model we selected is m_2: $\text{logit}(ch1) = \beta_0 + \beta_1 EV_1 - \beta_2 EV_2 + \beta_3 R_1 - \beta_4 R_2$, yielding the highest choice prediction accuracy of $82.87 \pm 0.25\%$ (mean $\pm$ s.e.m., $n = 108$ sessions; 83.87 ± 0.46 or subject 1, $82.21 \pm 0.28\%$ for subject 2; Supp. Fig. S2).

3. Neural data analyses

We analyzed the spiking activity of neurons in dACC to compute the fraction of cells that show significant encoding of the SV. Following the SV computation (Methods 2.4), we run a time-resolved analysis of the spike rate $\eta_i(t)$ of ith trial at time t in 200 ms boxcar time windows, sliding at 10 ms offset bins during task time. We used a linear model of the spike rate defined as: $\eta_i(t) = \beta_0(t) + \beta_1(t) SV_{1,i} + \beta_2(t) SV_{2,i}$, and tracked the fraction of cells showing significant encoding of the two SVs. The significance of SV interaction slope is assessed to be different from zero via two-tailed F-statistics tests on the sum of squared errors (SSE) of the full model $SSE^F = \sum_{i=1}^{n} \left|\eta_i - \left(\beta_0 + \hat{\beta}_1 SV_{1,i} + \hat{\beta}_2 SV_{2,i}\right)\right|^2$, and of the reduced model $SSE_{j=1,2}^R = \sum_{i=1}^{n} \left|\eta_i - \left(\beta_0 + \sum_{k=1,2;\, k\neq j} \hat{\beta}_k SV_{k,i}\right)\right|^2$, including only $\hat{\beta}_2$ for the assessment of $\hat{\beta}_1$, and, respectively, only $\hat{\beta}_1$ for the assessment of $\hat{\beta}_2$. The F-value for $\beta_{j=1,2}(t)$ is defined as $F_{j=1,2}^* = \left(\left(SSE_j^R - SSE^F\right)/1\right) / \left(SSE^F/(n-2)\right)$, with n being the number of trials observations. The p-values are computed by comparing empirical F-values with the F-distribution. The significance of the fractions of cells showing significant encoding is further assessed via the comparison of the significant fractions of cells with surrogate results run over $n = 1000$ trial order randomizations, independently generated

across variables. Lastly, consecutive runs of significant bins are assessed in length via run-length, cluster-based significance tests[20]. Note that difficulty and accumulated tokens count are not included among predictors, since we stratify the neural analyses using such variables.

3.1 Encoding of SV and accumulated reward

We further split the data based on the number of tokens accumulated on trials before the current trial, in the subsets 'Low' ($ATC = [0, 1]$ tokens) and 'High' ($ATC = [2, 5]$ tokens) accumulated tokens count. This criterion for binning in 'Low' and 'High' ATC is motivated by the interest in maximizing the number of trials used for spike count regression, and by considering that the number of tokens accumulated ATC is discrete, and its distribution shows many $ATC = 0$ entries. Since the number of entries could vary between the two subsets ($n = 225.55 \pm 4.28$ trials mean ± s.e.m. across sessions, approximately $45.18\% \pm 0.25\%$ of the total for 'Low' ATC, and $n = 274.15 \pm 5.25$ trials, approximately $54.82\% \pm 0.25\%$ of the total for 'High' ATC), the analysis is performed by randomly subsampling 'High' ATC trials in $n = 10$ randomly assigned subsets of size equal to the size of the 'Low' ATC trials subsets. Fractions of significant cells resulting from the $n = 10$ subsets were averaged across subsets before the statistical assignment. The fractions of bins showing significant encoding were averaged in time for each cell, then the average fractions of bins were compared via one-tailed, paired signed-rank tests, testing the hypothesis that in each cell the fraction of significant bins was higher for 'High' than for 'Low' ATC. As for the difficulty, also we checked whether the SV showed offset means and heteroskedasticity (Supp. Fig. S4-S5). The same analyses are also repeated on z-scored SVs, further checking that they lead to identical fractions of significant cells as for SVs prior to z-scoring (Supplementary Material).

3.2 Encoding of SV and difficulty

Following the same method as the one defined to compute the fraction of significant cells, we compute the fraction of significant cells in trial subsets based on their difficulty, i.e. we defined $\Delta_{SV} = |SV_1 - SV_2|$ term, computed its median value $= 1.9$ to distinguish Easy ($\Delta_{SV} < 1.9$) and Hard ($\Delta_{SV} \geq 1.9$) trials, and re-run in the neural encoding analyses. Note that for behavioral analyses we set difficulty based on Δ_{EV} (Methods 2.2). By using this difficulty assignment criteria, we could split trial pools in equally sized 'Hard' and 'Easy' subsets ($n = 250.91 \pm 8.18$ trials, mean ± s.e.m. across sessions, approximately $50.02\% \pm 1.23\%$ of the total for 'Hard', and $n = 250.90 \pm 8.10$ trials, approximately $49.98\% \pm 1.23\%$ of the total for 'Easy'). The fractions of bins showing significant encoding of SV_1 or SV_2 were averaged in time for each cell, then the average fractions of bins were compared via one-tailed, paired signed-rank tests, testing the hypothesis that in each cell the fraction of significant bins was higher for 'Easy' than for 'Hard' difficulty trials. Since the SV data show offset in mean and heteroskedasticity in Easy and Hard trial pools (Supp. Fig. S2-S3), we also run the same analyses on z-scored versions of the independent variables (SV_1, SV_2). It is straightforward to show that the sum of squared errors SSE^F, SSE_j^R, $j = 1,2$, and the coefficient of determination R^2 for z-scored independent data match the SSE over the same data prior to z-scoring. Thus the F^* values assessment provides numerical match, leading to identical results in terms of fractions of significant cells encoding the two values SV_1, SV_2 (Supplementary Material).

Data & code availability

Data & code is available at the DOI: 10.12751/g-node.1kkrw6.

References

1. Kahneman, D. & Tversky, A. *Prospect Theory: An Analysis of Decision under Risk*. vol. 47 https://about.jstor.org/terms (1979).
2. Anke Braun, X., Urai, A. E., Tobias, X. & Donner, H. Adaptive History Biases Result from Confidence-Weighted Accumulation of past Choices. *Journal of Neuroscience* **38**, (2018).
3. Nogueira, R. *et al.* Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. *Nat Commun* **8**, (2017).
4. Mochol, G., Kiani, R. & Moreno-Bote, R. Prefrontal cortex represents heuristics that shape choice bias and its integration into future behavior. *Current Biology* **31**, 1234-1244.e6 (2021).
5. Hermoso-Mendizabal, A. *et al.* Response outcomes gate the impact of expectations on perceptual decisions. *Nat Commun* **11**, (2020).
6. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).
7. Ebitz, R. B. & Hayden, B. Y. Dorsal anterior cingulate: A Rorschach test for cognitive neuroscience. *Nature Neuroscience* vol. 19 1278–1279 Preprint at https://doi.org/10.1038/nn.4387 (2016).
8. Haber, S. N. & Behrens, T. E. J. The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. *Neuron* vol. 83 1019–1039 Preprint at https://doi.org/10.1016/j.neuron.2014.08.031 (2014).
9. Heilbronner, S. R. & Hayden, B. Y. Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu Rev Neurosci* **39**, 149–170 (2016).
10. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* **10**, 1625–1633 (2007).
11. Camille, N., Griffiths, C. A., Vo, K., Fellows, L. K. & Kable, J. W. Ventromedial frontal lobe damage disrupts value maximization in humans. *Journal of Neuroscience* **31**, 7527–7532 (2011).
12. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci* **11**, 389–397 (2008).
13. Rushworth, M. F. S., Noonan, M. A. P., Boorman, E. D., Walton, M. E. & Behrens, T. E. Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* vol. 70 1054–1069 Preprint at https://doi.org/10.1016/j.neuron.2011.05.014 (2011).
14. Kennerley, S. W., Behrens, T. E. J. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* **14**, 1581–1589 (2011).
15. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
16. Wallis, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci* **30**, 31–56 (2007).
17. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci* **19**, 973–980 (2016).
18. Padoa-Schioppa, C. Neuronal origins of choice variability in economic decisions. *Neuron* **80**, 1322–1336 (2013).
19. Strait, C. E., Blanchard, T. C. & Hayden, B. Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* **82**, 1357–1366 (2014).
20. Ferro, D., Cash-Padgett, T., Wang, M. Z., Hayden, B. Y. & Moreno-Bote, R. Gaze-centered gating, reactivation, and reevaluation of economic value in orbitofrontal cortex. *Nat Commun* **15**, (2024).
21. Maisson, D. J. N. *et al.* Choice-relevant information transformation along a ventrodorsal axis in the medial prefrontal cortex. *Nat Commun* **12**, (2021).
22. Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S. & Cohen, J. D. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature 1999 402:6758* **402**, 179–181 (1999).

23. Van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A. & Carter, C. S. Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage* **14**, 1302–1308 (2001).
24. Botvinick, M. M., Cohen, J. D. & Carter, C. S. Conflict monitoring and anterior cingulate cortex: An update. *Trends Cogn Sci* **8**, 539–546 (2004).
25. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
26. Hadland, K. A., Rushworth, M. F. S., Gaffan, D. & Passingham, R. E. The anterior cingulate and reward-guided selection of actions. *J Neurophysiol* **89**, 1161–1164 (2003).
27. Vassena, E., Holroyd, C. B. & Alexander, W. H. Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Front Neurosci* **11**, (2017).
28. Kurniawan, I. T., Guitart-Masip, M., Dayan, P. & Dolan, R. J. Effort and valuation in the brain: The effects of anticipation and execution. *Journal of Neuroscience* **33**, 6160–6169 (2013).
29. Goh, A. X. A., Bennett, D., Bode, S. & Chong, T. T. J. Neurocomputational mechanisms underlying the subjective value of information. *Commun Biol* **4**, (2021).
30. Amiez, C., Joseph, J. P. & Procyk, E. Anterior cingulate error-related activity is modulated by predicted reward. *European Journal of Neuroscience* **21**, 3447–3452 (2005).
31. Vassena, E., Holroyd, C. B. & Alexander, W. H. Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Front Neurosci* **11**, (2017).
32. Aarts, E. & Roelofs, A. *Attentional Control in Anterior Cingulate Cortex Based on Probabilistic Cueing*. http://mitprc.silverchair.com/jocn/article-pdf/23/3/716/1774802/jocn.2010.21435.pdf (2010).
33. Strait, C. E. *et al.* Neuronal selectivity for spatial positions of offers and choices in five reward regions. *J Neurophysiol* **115**, 1098–1111 (2016).
34. Azab, H. & Hayden, B. Y. Correlates of decisional dynamics in the dorsal anterior cingulate cortex. *PLoS Biol* **15**, (2017).
35. Azab, H. & Hayden, B. Y. Correlates of economic decisions in the dorsal and subgenual anterior cingulate cortices. *European Journal of Neuroscience* **47**, 979–993 (2018).
36. Vassena, E., Deraeve, J. & Alexander, W. H. Surprise, value and control in anterior cingulate cortex during speeded decision-making. *Nat Hum Behav* **4**, 412–422 (2020).
37. Blanchard, T. C. & Hayden, B. Y. Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. *Journal of Neuroscience* **34**, 646–655 (2014).
38. Shidara, M. & Richmond, B. J. Anterior Cingulate: Single Neuronal Signals Related to Degree of Reward Expectancy. *Science (1979)* **296**, 1709–1711 (2002).
39. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Fictive reward signals in the anterior cingulate cortex. *Science (1979)* **324**, 948–950 (2009).
40. Kerns, J. G. *et al.* Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science (1979)* **303**, 1023–1026 (2004).
41. Hayden, B. Y., Heilbronner, S. R., Pearson, J. M. & Platt, M. L. Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience* **31**, 4178–4187 (2011).
42. Tversky, A. & Kahneman, D. Judgment under Uncertainty: Heuristics and Biases Biases in judgments reveal some heuristics of thinking under uncertainty. *Science (1979)* **185**, 1124–1131 (1974).
43. Tversky, A. & Kahneman, D. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *J Risk Uncertain* **5**, 297–323 (1992).
44. O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* **4**, 95–102 (2001).
45. Tobler, P. N., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J Neurophysiol* **97**, 1621–1632 (2007).
46. Lee, D., Seo, H. & Jung, M. W. Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience* vol. 35 287–308 Preprint at https://doi.org/10.1146/annurev-neuro-062111-150512 (2012).
47. Meder, D., Herz, D. M., Rowe, J. B., Lehéricy, S. & Siebner, H. R. The role of dopamine in the brain - lessons learned from Parkinson's disease. *NeuroImage* vol. 190 79–93 Preprint at https://doi.org/10.1016/j.neuroimage.2018.11.021 (2019).
48. Chakroun, K. *et al.* Dopamine regulates decision thresholds in human reinforcement learning in males. *Nat Commun* **14**, (2023).
49. Berridge, K. C. & Robinson, T. E. Parsing reward. *Trends Neurosci* **26**, 507–513 (2003).

50. Glimcher, P. W. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proc Natl Acad Sci U S A* **108**, 15647–15654 (2011).
51. Westbrook, A. *et al.* Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science (1979)* (2020).
52. Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J. & Rushworth, M. F. S. Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* **9**, 940–947 (2006).
53. Morcos, A. S. & Harvey, C. D. History-dependent variability in population dynamics during evidence accumulation in cortex. *Nat Neurosci* **19**, 1672–1681 (2016).
54. Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M. & Gardner, J. L. Adaptable history biases in human perceptual decisions. *Proc Natl Acad Sci U S A* **113**, E3548–E3557 (2016).

Acknowledgements

This project was supported by grants funded by the Spanish Ministry of Science, Innovation and Universities (MICIU/AEI/10.13039/501100011033) and by "FEDER A way of making Europe" (ref: PID2023-146524NB), and by ICREA ACADÈMIA (2022) funded by the Catalan Institution for Research and Advanced Studies to R.M.B.

Author contributions

D.F., G.M., R.M.B. conceptualized and designed the analyses. B.H. ideated the task. D.F., G.M. analyzed the data. B.H., H.A. collected the data. D.F., G. M., B.H., R.M.B. wrote the manuscript.

Declaration of interests

The authors declare no competing interests.